

# Semana 9: DISEÑO ESTADÍSTICO DE EXPERIMENTOS

## Table of contents

<b>1</b>	<b>Diseño Completamente Aleatorio con efectos fijos (Diseño unifactorial de efectos fijos)</b>	<b>3</b>
<b>2</b>	<b>Estudio de la Idoneidad del modelo</b>	<b>12</b>
2.1	Hipótesis de normalidad . . . . .	12
2.2	Hipótesis de homocedasticidad . . . . .	14
2.3	Hipótesis de independencia . . . . .	14
<b>3</b>	<b>Comparaciones múltiples</b>	<b>15</b>
<b>4</b>	<b>Diseño Unifactorial de efectos aleatorios</b>	<b>16</b>
<b>5</b>	<b>Diseño en Bloques Completos Aleatorizados</b>	<b>20</b>
5.1	Estudio de la Idoneidad del modelo . . . . .	23
5.1.1	Hipótesis de aditividad entre los bloques y tratamientos . . . . .	23
5.1.2	Hipótesis de Normalidad . . . . .	24
5.1.3	Hipótesis de Homogeneidad de Varianzas . . . . .	25
5.1.4	Hipótesis de Independencia . . . . .	25
5.2	Comparaciones múltiples . . . . .	26
<b>6</b>	<b>Diseño en bloques Incompletos Aleatorizados</b>	<b>29</b>
<b>7</b>	<b>Diseño en Cuadrados Latinos</b>	<b>34</b>
<b>8</b>	<b>Diseño en Cuadrados Greco-Latinos</b>	<b>36</b>
<b>9</b>	<b>Diseño en Cuadrados de Youden</b>	<b>39</b>

<b>10 Diseños Factoriales</b>	<b>43</b>
10.1 Diseños factoriales con dos factores . . . . .	43
10.1.1 El modelo sin replicación . . . . .	43
10.1.2 El modelo con replicación . . . . .	46
10.2 Diseños factoriales con tres factores . . . . .	49
10.2.1 El modelo sin replicación . . . . .	49
10.3 Diseño factorial de tres factores con replicación . . . . .	53
<b>11 Ejercicios Guiados</b>	<b>57</b>
<b>12 Ejercicios Propuestos</b>	<b>75</b>

# 1 Diseño Completamente Aleatorio con efectos fijos (Diseño unifactorial de efectos fijos)

## Supuesto práctico 1

La contaminación es uno de los problemas ambientales más importantes que afectan a nuestro mundo. En las grandes ciudades, la contaminación del aire se debe a los escapes de gases de los motores de explosión, a los aparatos domésticos de la calefacción, a las industrias,... El aire contaminado nos afecta en nuestro vivir diario, manifestándose de diferentes formas en nuestro organismo. Con objeto de comprobar la contaminación del aire en una determinada ciudad, se ha realizado un estudio en el que se han analizado las concentraciones de monóxido de carbono (CO) durante cinco días de la semana (lunes, martes, miércoles, jueves y viernes).

- Variable respuesta: Concentración de CO.
- Factor: Día de la semana que tiene cinco niveles. Es un factor de efectos fijos ya que viene decidido qué niveles concretos se van a utilizar (5 días de la semana).
- Modelo equilibrado: Los niveles de los factores tienen el mismo número de elementos (8 elementos).
- Tamaño del experimento: Número total de observaciones, en este caso 40 unidades experimentales.

El problema planteado se modeliza a través de un **diseño unifactorial totalmente aleatorizado de efectos fijos equilibrado**.

Para realizar este supuesto en R debemos introducir primero los datos de forma correcta. Podemos realizarlo directamente en R de forma manual o introducirlos previamente en un archivo de texto o Excel y leerlos en R.

```
contaminacion <- read.table('supuesto1.txt', header = TRUE)
contaminacion
```

	Concentracion	Dia
1	420	lunes
2	390	lunes
3	480	lunes
4	430	lunes
5	440	lunes
6	324	lunes
7	450	lunes
8	460	lunes
9	450	martes

10	390	martes
11	430	martes
12	521	martes
13	320	martes
14	360	martes
15	342	martes
16	423	martes
17	355	miercoles
18	462	miercoles
19	286	miercoles
20	238	miercoles
21	344	miercoles
22	423	miercoles
23	123	miercoles
24	196	miercoles
25	321	jueves
26	254	jueves
27	412	jueves
28	368	jueves
29	340	jueves
30	258	jueves
31	433	jueves
32	489	jueves
33	238	viernes
34	255	viernes
35	366	viernes
36	389	viernes
37	198	viernes
38	256	viernes
39	248	viernes
40	324	viernes

1. Transformar la variable referente a los niveles del factor fijo como factor

```
contaminacion$dia<-factor(contaminacion$Dia)
contaminacion$dia
```

```
[1] lunes    lunes    lunes    lunes    lunes    lunes    lunes
[8] lunes    martes   martes   martes   martes   martes   martes
[15] martes   martes   miercoles miercoles miercoles miercoles miercoles
[22] miercoles miercoles miercoles jueves    jueves    jueves    jueves
[29] jueves    jueves    jueves    jueves    viernes   viernes   viernes
```

```
[36] viernes  viernes  viernes  viernes  viernes
Levels: jueves lunes martes miercoles viernes
```

Para calcular la tabla ANOVA primero hacemos uso de la función “aov” de la siguiente forma:

```
mod<-aov(Concentracion~Dia,data = contaminacion);mod
```

Call:

```
aov(formula = Concentracion ~ Dia, data = contaminacion)
```

Terms:

	Dia	Residuals
Sum of Squares	119484.4	218948.8
Deg. of Freedom	4	35

Residual standard error: 79.09285

Estimated effects may be unbalanced

donde:

- Concentracion = nombre de la columna de las observaciones.
- Dia = nombre de la columna en la que están representados los tratamientos.
- data= data.frame en el que están guardados los datos.

Se puede mostrar un resumen de los resultados con la función “summary” (verdadera tabla ANOVA)

```
summary(mod)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Dia	4	119484	29871	4.775	0.00352 **
Residuals	35	218949	6256		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Si el valor de F es mayor que uno quiere decir que hay un efecto positivo del factor día. Se observa que el P-valor (Sig.) tiene un valor de 0.003524, que es menor que el nivel de significación 0.05. Por lo tanto, hemos comprobado estadísticamente que estos cinco grupos son distintos. Es decir, existen diferencias significativas en las concentraciones medias de monóxido de carbono entre los cinco días de la semana. Por lo tanto no se puede rechazar la

hipótesis alternativa que dice que al menos dos grupos son diferentes, pero ¿Cuáles son esos grupos? ¿Los cinco grupos son distintos o sólo alguno de ellos? Pregunta que resolveremos más adelante mediante los contrastes de comparaciones múltiples.

2. En la expresión del comando “aov” indicar el factor

```
mod1<-aov(Concentracion~factor(Dia),data = contaminacion);mod1
```

Call:

```
aov(formula = Concentracion ~ factor(Dia), data = contaminacion)
```

Terms:

	factor(Dia)	Residuals
Sum of Squares	119484.4	218948.8
Deg. of Freedom	4	35

Residual standard error: 79.09285

Estimated effects may be unbalanced

```
summary(mod1)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
factor(Dia)	4	119484	29871	4.775	0.00352 **
Residuals	35	218949	6256		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

También se puede utilizar el comando “anova” y no es necesario el comando “summary”

```
mod2 <- anova (lm (Concentracion ~ factor (Dia), data = contaminacion));mod2
```

Analysis of Variance Table

Response: Concentracion

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
factor(Dia)	4	119484	29871.1	4.775	0.003518 **
Residuals	35	218949	6255.7		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Los datos pueden venir dados en diferentes formatos:

1. Caso en el que los datos se muestran de forma que se analiza la contaminación con cada uno de los días de la semana (de lunes a viernes). Como se muestra a continuación

```
contaminacion <- read.table('supuesto1-1.txt', header = TRUE)
contaminacion
```

	Lunes	Martes	Miercoles	Jueves	Viernes
1	420	450	355	321	238
2	390	390	462	254	255
3	480	430	286	412	366
4	430	521	238	368	389
5	440	320	344	340	198
6	324	360	423	258	256
7	450	342	123	433	248
8	460	423	196	489	324

En primer lugar apilaremos las columnas, para ello utilizamos el comando “stack” de la siguiente forma

```
trats<-stack(contaminacion);trats
```

	values	ind
1	420	Lunes
2	390	Lunes
3	480	Lunes
4	430	Lunes
5	440	Lunes
6	324	Lunes
7	450	Lunes
8	460	Lunes
9	450	Martes
10	390	Martes
11	430	Martes
12	521	Martes
13	320	Martes
14	360	Martes
15	342	Martes
16	423	Martes
17	355	Miercoles

18	462	Miercoles
19	286	Miercoles
20	238	Miercoles
21	344	Miercoles
22	423	Miercoles
23	123	Miercoles
24	196	Miercoles
25	321	Jueves
26	254	Jueves
27	412	Jueves
28	368	Jueves
29	340	Jueves
30	258	Jueves
31	433	Jueves
32	489	Jueves
33	238	Viernes
34	255	Viernes
35	366	Viernes
36	389	Viernes
37	198	Viernes
38	256	Viernes
39	248	Viernes
40	324	Viernes

Nos muestra dos columnas:

- La primera columna: values nos muestra los valores de la variable respuesta. En este caso la contaminación
- La segunda columna: ind nos muestra los diferentes tratamientos

Podemos realizar el Análisis de la varianza utilizando el comando **anova**

```
anova(lm(values ~ ind, data = trats))
```

Analysis of Variance Table

Response: values

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ind	4	119484	29871.1	4.775	0.003518 **
Residuals	35	218949	6255.7		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1



2. Los datos vienen dados de la siguiente forma:

Lunes: 420, 390, 480, 430, 440, 324, 450, 460

Martes: 450, 390, 430, 521, 320, 360, 342, 423

Miércoles: 355, 462, 286, 238, 344, 423, 123, 196

Jueves: 321, 254, 412, 368, 340, 258, 433, 489

Viernes: 238, 255, 366, 389, 198, 256, 248, 324

Se crean cinco vectores, cada uno de ellos representando la contaminación con un tratamiento.

```
Lu= c(420, 390, 480, 430, 440, 324, 450, 460)
Ma = c(450, 390, 430, 521, 320, 360, 342, 423)
Mi= c(355, 462, 286, 238, 344, 423, 123, 196)
Ju=c(321, 254, 412, 368, 340, 258, 433, 489)
Vi= c(238, 255, 366, 389, 198, 256, 248, 324)

Lu;Ma;Mi;Ju;Vi
```

```
[1] 420 390 480 430 440 324 450 460
```

```
[1] 450 390 430 521 320 360 342 423
```

```
[1] 355 462 286 238 344 423 123 196
```

```
[1] 321 254 412 368 340 258 433 489
```

```
[1] 238 255 366 389 198 256 248 324
```

Acontinuación creamos un data.frame para poder resolver el ANOVA

```
datos = data.frame(Lu, Ma, Mi, Ju,Vi)
datos
```

	Lu	Ma	Mi	Ju	Vi
1	420	450	355	321	238
2	390	390	462	254	255
3	480	430	286	412	366
4	430	521	238	368	389
5	440	320	344	340	198
6	324	360	423	258	256
7	450	342	123	433	248
8	460	423	196	489	324

De esta forma hemos creado una nueva base de datos que hemos llamado “**datos**”. Para resolver el ANOVA tenemos primero que apilar las columnas con el comando “**stack**”

```
datos1 = stack(datos)
datos1
```

	values	ind
1	420	Lu
2	390	Lu
3	480	Lu
4	430	Lu
5	440	Lu
6	324	Lu
7	450	Lu
8	460	Lu
9	450	Ma
10	390	Ma
11	430	Ma
12	521	Ma
13	320	Ma
14	360	Ma
15	342	Ma
16	423	Ma
17	355	Mi
18	462	Mi
19	286	Mi
20	238	Mi
21	344	Mi
22	423	Mi
23	123	Mi
24	196	Mi

```

25    321  Ju
26    254  Ju
27    412  Ju
28    368  Ju
29    340  Ju
30    258  Ju
31    433  Ju
32    489  Ju
33    238  Vi
34    255  Vi
35    366  Vi
36    389  Vi
37    198  Vi
38    256  Vi
39    248  Vi
40    324  Vi

```

```

# Resolvemos el ANOVA como en el caso anterior

anova(lm(values ~ ind, data = datos1))

```

#### Analysis of Variance Table

Response: values

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
ind	4	119484	29871.1	4.775	0.003518 **
Residuals	35	218949	6255.7		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

3. Los datos se muestran en un solo vector que tiene todos los datos de la contaminación tanto si se ha medido el lunes, el martes, el miércoles, el jueves o el viernes

```

contaminacion = c(Lu, Ma, Mi, Ju, Vi)
contaminacion

```

```

[1] 420 390 480 430 440 324 450 460 450 390 430 521 320 360 342 423 355 462 286
[20] 238 344 423 123 196 321 254 412 368 340 258 433 489 238 255 366 389 198 256
[39] 248 324

```

```
# Comando length
length(contaminacion)
```

```
[1] 40
```

Para realizar el ANOVA, ya tenemos los datos de la variable respuesta y a continuación tenemos que crear el factor tratamiento, para ello vamos a utilizar la función `gl`, y le decimos que nos genere 5 niveles que son los cinco tratamientos, cada uno repetido 8 veces con un total de 40 datos y para identificar que nivel es cada uno, creamos las etiquetas L, M, Mi, J y V

```
trat = gl(5,8,40, labels= c('L', 'M', 'Mi', 'j', 'V'));trat
```

```
[1] L L L L L L L L M M M M M M M M Mi Mi Mi Mi Mi Mi Mi Mi j
[26] j j j j j j j j V V V V V V V V V
Levels: L M Mi j V
```

```
anova(lm(contaminacion~trat))
```

Analysis of Variance Table

Response: `contaminacion`

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
trat	4	119484	29871.1	4.775	0.003518 **
Residuals	35	218949	6255.7		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

El modelo que hemos propuesto hay que validarlo, para ello hay que comprobar si se verifican las hipótesis básicas del modelo, es decir, si las perturbaciones son variables aleatorias independientes con distribución normal de media 0 y varianza constante (homocedasticidad).

## 2 Estudio de la Idoneidad del modelo

### 2.1 Hipótesis de normalidad

En primer lugar, analizamos la normalidad de las concentraciones y continuamos con el análisis de la normalidad de los residuos.

Para analizar la normalidad de las concentraciones utilizamos el test de Shapiro-Wilks

```
shapiro.test(mod$residuals)
```

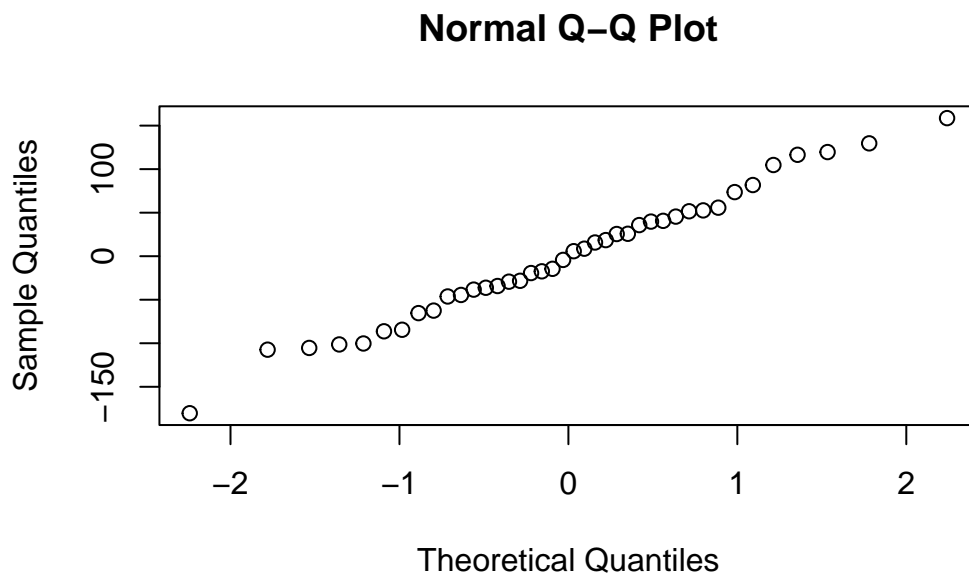
Shapiro-Wilk normality test

```
data:  mod$residuals  
W = 0.98933, p-value = 0.9654
```

Observamos el contraste de Shapiro-Wilk que es adecuado cuando las muestras son pequeñas ( $n < 50$ ) y es una alternativa más potente que el test de Kolmogorov-Smirnov. El p-valor es mayor que el nivel de significación del 5%, concluyendo que las muestras de las concentraciones se distribuyen de forma normal en cada día de la semana.

Podemos verlo también gráficamente con la orden “qqnorm”

```
qqnorm (mod$residuals)
```



Podemos apreciar en este gráfico que los puntos aparecen próximos a la línea diagonal. Esta gráfica no muestra una desviación marcada de la normalidad.

## 2.2 Hipótesis de homocedasticidad

Para comprobar la hipótesis de igualdad entre las varianzas del factor utilizamos el Test de Barlett.

```
contaminacion <- read.table('supuesto1.txt', header = TRUE)
contaminacion$dia<-factor(contaminacion$Dia)
bartlett.test(contaminacion$Concentracion, contaminacion$Dia)
```

```
Bartlett test of homogeneity of variances
```

```
data:  contaminacion$Concentracion and contaminacion$Dia
Bartlett's K-squared = 5.4942, df = 4, p-value = 0.2402
```

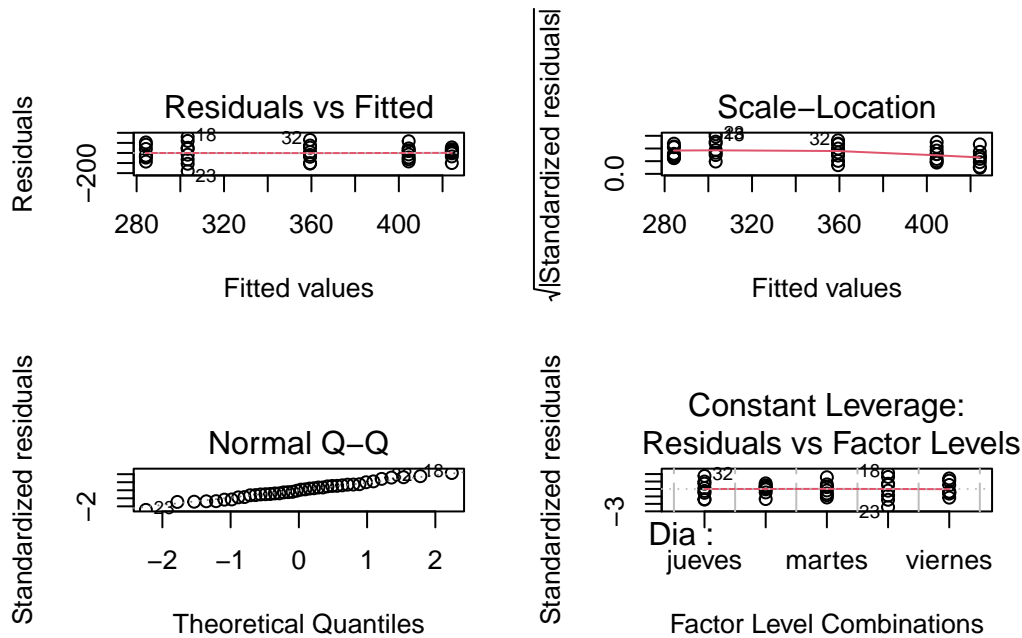
El p-valor es del 0.2402 que al ser mayor del nivel significación usual del 5% no podemos rechazar la hipótesis de igualdad de varianzas, es decir, se acepta la igualdad de varianzas en el factor.

## 2.3 Hipótesis de independencia

Para comprobar que se satisface el supuesto de independencia entre los residuos analizamos el gráfico de los residuos frente a los valores pronosticados o predichos por el modelo. El empleo de este gráfico es útil puesto que la presencia de alguna tendencia en el mismo puede ser indicio de una violación de dicha hipótesis. En R obtenemos varios gráficos a la vez que están incluidos en la estimación del modelo.

Para verlos de forma correcta hacemos uso de las siguientes órdenes:

```
layout(matrix(c(1,2,3,4),2,2)) # para que salgan en la misma pantalla
plot(mod)
```



En la Figura 5 se muestran cuatro gráficos, en el primero de ellos que se representan los residuos en el eje de ordenadas y los valores pronosticados en el eje de abscisas. No observamos, en dicho gráfico, ninguna tendencia sistemática que haga sospechar del incumplimiento de la suposición de independencia.

Anteriormente, hemos comprobado estadísticamente que estos cinco grupos son distintos. Es decir no se puede rechazar la hipótesis alternativa que dice que al menos dos grupos son diferentes, pero ¿Cuáles son esos grupos? ¿Los cinco grupos son distintos o sólo alguno de ellos? Pregunta que resolveremos más adelante mediante los contrastes de comparaciones múltiples.

### 3 Comparaciones múltiples

Para saber entre que parejas de días las diferencias entre concentraciones medias de CO son significativas aplicamos la prueba **Post-hoc de Tukey**

```
mod.tukey<- TukeyHSD(mod, ordered = TRUE)
mod.tukey
```

```
Tukey multiple comparisons of means
 95% family-wise confidence level
```

factor levels have been ordered

Fit: aov(formula = Concentracion ~ Dia, data = contaminacion)

\$Dia

	diff	lwr	upr	p adj
miercoles-viernes	19.125	-94.573356	132.8234	0.9883811
jueves-viernes	75.125	-38.573356	188.8234	0.3363682
martes-viernes	120.250	6.551644	233.9484	0.0337946
lunes-viernes	140.000	26.301644	253.6984	0.0095230
jueves-miercoles	56.000	-57.698356	169.6984	0.6217479
martes-miercoles	101.125	-12.573356	214.8234	0.1010091
lunes-miercoles	120.875	7.176644	234.5734	0.0325284
martes-jueves	45.125	-68.573356	158.8234	0.7837763
lunes-jueves	64.875	-48.823356	178.5734	0.4826413
lunes-martes	19.750	-93.948356	133.4484	0.9868896

Esta salida nos muestra los intervalos de confianza simultáneos construidos por el método de Tukey. En la tabla se muestra un resumen de las comparaciones de cada tratamiento con los restantes. Es decir, aparecen comparadas dos a dos las cinco medias de los tratamientos.

## 4 Diseño Unifactorial de efectos aleatorios

### Supuesto práctico 2

Los medios de cultivo bacteriológico en los laboratorios de los hospitales proceden de diversos fabricantes. Se sospecha que la calidad de estos medios de cultivo varía de un fabricante a otro. Para comprobar esta teoría, se hace una lista de fabricantes de un medio de cultivo concreto, se seleccionan aleatoriamente los nombres de cinco de los que aparecen en la lista y se comparan las muestras de los instrumentos procedentes de éstos. La comprobación se realiza colocando sobre una placa dos dosis, en gotas, de una suspensión medida de un microorganismo clásico, *Escherichia coli*, dejando al cultivo crecer durante veinticuatro horas, y determinando después el número de colonias (en millares) del microorganismo que aparecen al final del período. Se quiere comprobar si la calidad del instrumental difiere entre fabricantes.

Supuestos del modelo

- Las cinco muestras representan muestras aleatorias independientes extraídas de I poblaciones seleccionadas aleatoriamente de un conjunto mayor de poblaciones.
- Todas las poblaciones del conjunto más amplio tienen distribución Normal, de modo que cada una de las 5 poblaciones muestreadas se distribuyen según una Normal



- Todas las poblaciones del conjunto más amplio tienen la misma varianza, y por lo tanto, cada una de las 5 poblaciones muestreadas tiene también varianza  $\sigma^2$ .
- Las variables  $\tau_i$  son variables aleatorias normales independientes, cada una con media 0 y varianza común  $\sigma_\tau^2$ .

### En el supuesto práctico 2:

- **Variable respuesta: Calidad\_Instrumental**
- **Factor: Fabricante.** Es un factor de efectos aleatorios, se han elegido aleatoriamente a cinco fabricantes, que constituyen únicamente una muestra de todos los fabricantes y el propósito no es comparar estos cinco fabricantes sino contrastar el supuesto general de que la calidad del instrumental difiere entre fabricantes.
- **Modelo equilibrado:** Los niveles de los factores tienen el mismo número de elementos (9 elementos).
- **Tamaño del experimento:** Número total de observaciones, en este caso 45 unidades experimentales.

El problema planteado se modeliza a través de un diseño unifactorial totalmente aleatorizado de efectos aleatorios equilibrado.

```
bacterias<-read.table('supuesto2.txt', header = TRUE)
bacterias
```

	Calidad	Fabricante
1	120	1
2	240	2
3	240	3
4	300	4
5	300	5
6	240	1
7	360	2
8	270	3
9	240	4
10	360	5
11	300	1
12	180	2
13	300	3
14	300	4
15	240	5
16	360	1
17	180	2

18	360	3
19	360	4
20	360	5
21	240	1
22	300	2
23	360	3
24	360	4
25	360	5
26	180	1
27	240	2
28	300	3
29	360	4
30	360	5
31	144	1
32	360	2
33	360	3
34	360	4
35	360	5
36	300	1
37	360	2
38	360	3
39	360	4
40	300	5
41	240	1
42	360	2
43	300	3
44	300	4
45	360	5

Debemos transformar la variable referente a los niveles del factor fijo como factor para poder hacer los cálculos de forma adecuada

```
bacterias$Fabricante<- factor(bacterias$Fabricante)
bacterias$Fabricante
```

```
[1] 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3 4 5 1 2 3
[39] 4 5 1 2 3 4 5
Levels: 1 2 3 4 5
```

Para calcular la tabla ANOVA primero hacemos uso de la función “aov” de la siguiente forma:

```
mod <- aov(Calidad ~ Fabricante, data = bacterias)
mod
```

Call:

```
aov(formula = Calidad ~ Fabricante, data = bacterias)
```

Terms:

	Fabricante	Residuals
Sum of Squares	57363.2	144272.0
Deg. of Freedom	4	40

Residual standard error: 60.05664

Estimated effects may be unbalanced

donde:

- Calidad = nombre de la columna de las observaciones.
- Fabricante = nombre de la columna en la que están representados los tratamientos.
- data = data.frame en el que están guardados los datos.

y posteriormente mostramos un resumen de los resultados con la función “summary” (verdadera tabla ANOVA):

```
# TABLA ANOVA
summary(mod)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Fabricante	4	57363	14341	3.976	0.00827 **
Residuals	40	144272	3607		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Esta tabla muestra los resultados del contraste planteado. El valor del estadístico de contraste es igual a **3.976** que deja a la derecha un p-valor de **0.00827**, así que la respuesta dependerá del nivel de significación que se fije. Si fijamos un nivel de significación de 0.05 se concluye que hay evidencia suficiente para afirmar la existencia de alguna variabilidad entre la calidad del material de los diferentes fabricantes. Si fijamos un nivel de significación de 0.001, no podemos hacer tal afirmación.

## 5 Diseño en Bloques Completos Aleatorizados

### Supuesto práctico 3

Abeto blanco, Abeto del Pirineo, es un árbol de gran belleza por la elegancia de sus formas y el exquisito perfume balsámico que destilan sus hojas y cortezas. Destilando hojas y madera se obtiene aceite de trementina muy utilizado en medicina contra torceduras y contusiones. En estos últimos años se ha observado que la producción de semillas ha descendido y con objeto de conseguir buenas producciones se proponen tres tratamientos. Se observa que árboles diferentes tienen distintas características naturales de reproducción, este efecto de las diferencias entre los árboles se debe de controlar y este control se realiza mediante bloques. En el experimento se utilizan 10 abetos, dentro de cada abeto se seleccionan tres ramas semejantes. Cada rama recibe exactamente uno de los tres tratamientos que son asignados aleatoriamente. Constituyendo cada árbol un bloque completo. Los datos obtenidos se presentan en la siguiente tabla donde se muestra el número de semillas producidas por rama.

```
semillas<-read.table('supuesto3.txt', header = TRUE)
semillas
```

	y	Tratamiento	Abeto
1	7	1	1
2	9	2	1
3	10	3	1
4	8	1	2
5	9	2	2
6	10	3	2
7	9	1	3
8	9	2	3
9	12	3	3
10	10	1	4
11	9	2	4
12	12	3	4
13	11	1	5
14	12	2	5
15	14	3	5
16	8	1	6
17	10	2	6
18	9	3	6
19	7	1	7
20	8	2	7
21	7	3	7
22	8	1	8

23	8	2	8
24	7	3	8
25	7	1	9
26	9	2	9
27	10	3	9
28	8	1	10
29	9	2	10
30	10	3	10

A continuación debemos transformar tanto la columna de los tratamientos como la de los bloques en un factor para poder realizar los cálculos posteriores adecuadamente.

```
semillas$Tratamiento = factor(semillas$Tratamiento)
semillas$Tratamiento
```

```
[1] 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3 1 2 3
Levels: 1 2 3
```

```
semillas$Abeto = factor(semillas$Abeto)
semillas$Abeto
```

```
[1] 1 1 1 2 2 2 3 3 3 4 4 4 5 5 5 6 6 6 7 7 7 8 8 8 9
[26] 9 9 10 10 10
Levels: 1 2 3 4 5 6 7 8 9 10
```

```
# Calculando la tabla ANOVA
mod = aov(y ~ Tratamiento + Abeto, data = semillas);mod
```

Call:

```
aov(formula = y ~ Tratamiento + Abeto, data = semillas)
```

Terms:

	Tratamiento	Abeto	Residuals
Sum of Squares	16.2	54.8	15.8
Deg. of Freedom	2	9	18

Residual standard error: 0.936898

Estimated effects may be unbalanced

donde:

- **y** es el nombre de la columna de las observaciones
- **Tratamiento** es el nombre de la columna en la que están representados los tratamientos
- **Abeto** es el nombre de la columna en la que están representados los bloques
- **data = data.frame** en el que están guardados los datos

y a continuación mostramos un resumen de los resultados con la función “summary” (verdadera tabla ANOVA):

```
summary(mod)
```

```
              Df Sum Sq Mean Sq F value    Pr(>F)
Tratamiento    2   16.2    8.100    9.228 0.00174 **
Abeto           9   54.8    6.089    6.937 0.00026 ***
Residuals     18   15.8    0.878
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

La Tabla ANOVA, muestra que:

- El valor del estadístico de contraste de igualdad de bloques, **F = 6.937** deja a su derecha un p-valor menor que **0.001**, menor que el nivel de significación del 5%, por lo que se rechaza la Hipótesis nula de igualdad de bloques. La eficacia de este diseño depende de los efectos de los bloques. Un valor grande de F de los bloques (6.937) implica que el factor bloque tiene un efecto grande. En este caso el diseño es más eficaz que el diseño completamente aleatorizado ya que si el cuadrado medio entre bloques es grande (6.089), el término residual será mucho menor (0.878) y el contraste principal de las medias de los tratamientos será más sensible a las diferencias entre tratamientos. Por lo tanto la inclusión del factor bloque en el modelo es acertada. Así, la producción de semillas depende del abeto. Si los efectos de los bloques son muy pequeños, el análisis de bloque quizás no sea necesario y en caso extremo, cuando el valor de F de los bloques es próximo a 1, puede llegar a ser perjudicial, ya que el número de grados de libertad,  $(I-1)(J-1)$ , del denominador de la comparación de tratamientos es menor que el número de grados de libertad correspondiente,  $IJ-I$ , en el diseño completamente aleatorizado. Pero, ¿Cómo saber cuándo se puede prescindir de los bloques? La respuesta la tenemos en el valor de la F experimental de los bloques, se ha comprobado que si dicho valor es mayor que 3, no conviene prescindir de los bloques para efectuar los contrastes.
- El valor del estadístico de contraste de igualdad de tratamiento, **F = 9.228** deja a su derecha un p-valor de **0.002**, menor que el nivel de significación del 5%, por lo que se rechaza la Hipótesis nula de igualdad de tratamientos. Así, los tratamientos influyen en

el número de semillas. Es decir, existen diferencias significativas en el número de semillas entre los tres tratamientos.

El modelo que hemos propuesto hay que validarlo, para ello hay que comprobar si se verifican los cuatros supuestos expresados anteriormente.

## 5.1 Estudio de la Idoneidad del modelo

### 5.1.1 Hipótesis de aditividad entre los bloques y tratamientos

La interacción entre el factor bloque y los tratamientos vamos a estudiarla analíticamente mediante el Test de Interacción de un grado de Tukey

Para realizar este test en R tenemos que utilizar la library “daewr” y dentro de ella la función “Tukey1df”. De la siguiente forma:

Primero hay que instalar el paquete daewr

Para ello, seleccionar Paquetes/Instalar paquetes y de la lista escoger daewr. O bien utilizar la siguiente orden

```
#utils:::menuInstallPkgs()
```

Para realizar este contraste hay que utilizar la libray daewr, para ello realizamos la siguiente orden

```
library(daewr)
```

Registered S3 method overwritten by 'DoE.base':

```
method      from  
factorize.factor conf.design
```

```
Tukey1df(semillas)
```

Source	df	SS	MS	F	Pr>F
A	2	16.2	8.1		
B	9	54.8	6.0889		
Error	18	15.8	71.1		
NonAdditivity	1	3.5573	3.5573	4.94	0.0401
Residual	17	12.2427	0.7202		

Puesto que ( $Pr < F$ ), no hay suficiente evidencia para aceptar la hipotesis nula

### 5.1.2 Hipótesis de Normalidad

La normalidad las vamos a comprobar analíticamente y gráficamente.

Analíticamente mediante el contraste de Shapiro-Wilk que es adecuado cuando las muestras son pequeñas ( $n < 50$ )

```
shapiro.test(mod$residuals)
```

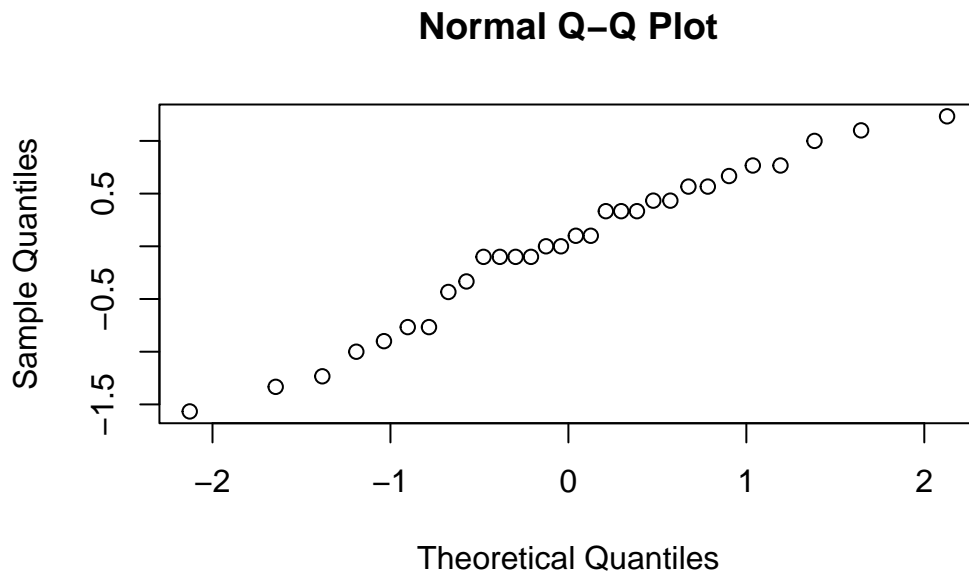
Shapiro-Wilk normality test

```
data:  mod$residuals  
W = 0.96415, p-value = 0.3935
```

Como podemos observar tenemos un p-valor de 0.3935 que aceptaría la hipótesis de normalidad por ser mayor al 5% (nivel de significación usual).

Gráficamente mediante el gráfico probabilístico normal. Para ello utilizamos la orden “qqnorm”

```
qqnorm(mod$residuals)
```





En esta gráfica observamos que prácticamente todos los puntos se encuentran sobre la diagonal por lo tanto podemos decir que no muestra una desviación marcada de la normalidad.

### 5.1.3 Hipótesis de Homogeneidad de Varianzas

Para comprobar la hipótesis de homocedasticidad utilizamos el Test de Barlett distinguiendo entre la igualdad entre varianzas del factor principal y la igualdad de varianzas del factor bloque.

En nuestro ejemplo, el test para igualdad de varianzas del factor principal sería:

```
bartlett.test(semillas$y,semillas$Tratamiento)
```

```
Bartlett test of homogeneity of variances
```

```
data:  semillas$y and semillas$Tratamiento  
Bartlett's K-squared = 4.1729, df = 2, p-value = 0.1241
```

El p-valor es del 0.1241 que al ser mayor del nivel significación usual del 5% no podemos rechazar la hipótesis de igualdad de varianzas en el factor principal.

De la misma manera procedemos para el factor bloque:

```
bartlett.test(semillas$y,semillas$Abeto)
```

```
Bartlett test of homogeneity of variances
```

```
data:  semillas$y and semillas$Abeto  
Bartlett's K-squared = 4.0723, df = 9, p-value = 0.9066
```

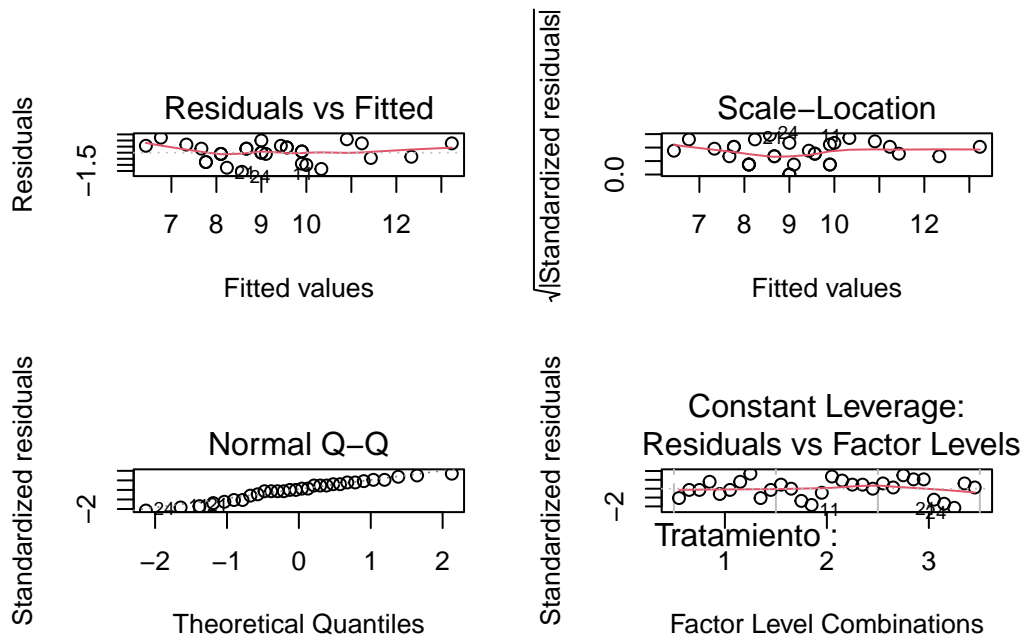
El p-valor es mayor que 0.05 por lo que no podemos rechazar la hipótesis de igualdad de varianzas en el factor bloque.

### 5.1.4 Hipótesis de Independencia

Comprobaremos si se satisface el supuesto de independencia entre los residuos. Para ello tenemos que representar un gráfico de los residuos tipificados frente a los pronosticados. En R obtenemos varios gráficos a la vez que están incluidos en la estimación del modelo.

Para verlos de forma correcta hacemos uso de las siguientes órdenes:

```
layout(matrix(c(1,2,3,4),2,2))
plot(mod)
```



Nos fijamos en el primer gráfico que representa los residuos frente a los valores ajustados y observamos que no hay ninguna tendencia sistemática. Concluimos que no hay sospechas para que se incumpla la hipótesis de independencia.

## 5.2 Comparaciones múltiples

Hemos probado anteriormente que se rechaza la Hipótesis nula de igualdad de tratamientos. Así, los tratamientos influyen en el número de semillas. Es decir, existen diferencias significativas en el número de semillas entre los tres tratamientos. Para saber entre que parejas de días estas diferencias son significativas aplicamos una prueba **Post-hoc**.

El contraste de Comparaciones múltiples que vamos a utilizar es el Test de Duncan. Para poder hacer uso de él en R tenemos que instalar en primer lugar el paquete “**agricolae**” y dentro de él la función “**duncan.test**”.

Destacar que este test hace las comparaciones especificándole si es para el factor principal o el factor bloque.

Comenzamos con el factor principal:

```
library(agricolae)
(duncan=duncan.test(mod,"Tratamiento",group=T))
```

```
$statistics
      MSError Df Mean      CV
0.8777778 18  9.2 10.18367

$parameters
      test      name.t ntr alpha
Duncan Tratamiento   3  0.05

$duncan
      Table CriticalRange
2 2.971152      0.8802727
3 3.117384      0.9235973

$means
      y      std  r Min Max  Q25 Q50  Q75
1  8.3 1.337494 10   7  11 7.25   8  8.75
2  9.2 1.135292 10   8  12 9.00   9  9.00
3 10.1 2.183270 10   7  14 9.25  10 11.50

$comparison
NULL

$groups
      y groups
3 10.1      a
2  9.2      b
1  8.3      c

attr(,"class")
[1] "group"
```

En el apartado “\$groups” concluimos que los tres tratamientos difieren significativamente entre sí.

Se observa que la concentración media del número de semillas es mayor con el Tratamiento3 (10.1) y menor con el Tratamiento1 (8.3).

Para el factor bloque:

```
(duncan=duncan.test(mod, "Abeto" , group = T))
```

```
$statistics
```

```
      MSError Df Mean      CV
0.8777778 18  9.2 10.18367
```

```
$parameters
```

```
test name.t ntr alpha
Duncan  Abeto 10  0.05
```

```
$duncan
```

```
      Table CriticalRange
2  2.971152      1.607151
3  3.117384      1.686250
4  3.209655      1.736161
5  3.273593      1.770746
6  3.320327      1.796026
7  3.355651      1.815133
8  3.382941      1.829895
9  3.404326      1.841462
10 3.421226      1.850604
```

```
$means
```

```
      y      std r Min Max  Q25 Q50  Q75
1  8.666667 1.5275252 3  7  10  8.0  9  9.5
10 9.000000 1.0000000 3  8  10  8.5  9  9.5
2  9.000000 1.0000000 3  8  10  8.5  9  9.5
3 10.000000 1.7320508 3  9  12  9.0  9 10.5
4 10.333333 1.5275252 3  9  12  9.5 10 11.0
5 12.333333 1.5275252 3 11  14 11.5 12 13.0
6  9.000000 1.0000000 3  8  10  8.5  9  9.5
7  7.333333 0.5773503 3  7  8  7.0  7  7.5
8  7.666667 0.5773503 3  7  8  7.5  8  8.0
9  8.666667 1.5275252 3  7  10  8.0  9  9.5
```

```
$comparison
```

```
NULL
```

```
$groups
```

```
      y groups
5 12.333333  a
4 10.333333  b
```

3	10.000000	b
10	9.000000	bc
2	9.000000	bc
6	9.000000	bc
1	8.666667	bc
9	8.666667	bc
8	7.666667	c
7	7.333333	c

```
attr(,"class")
[1] "group"
```

Se observa que la prueba de Duncan ha agrupado los abetos 7, 8, 1, 9, 2, 6 y 10 en un mismo grupo, 1, 9, 2, 6, 10, 3 y 4, en otro grupo y un tercer está formada únicamente por el Abeto5. Inmediatamente se ve que por ejemplo el Abeto5 difiere de todos los demás, siendo en este abeto donde se produce el mayor número de semillas (12.333) y el menor en el Abeto7 (7.333).

## 6 Diseño en bloques Incompletos Aleatorizados

### Supuesto práctico 4

Se realiza un estudio para comprobar la efectividad en el retraso del crecimiento de bacterias utilizando cuatro soluciones diferentes para lavar los envases de la leche. El análisis se realiza en el laboratorio y sólo se pueden realizar seis pruebas en un mismo día. Como los días son una fuente de variabilidad potencial, el investigador decide utilizar un diseño aleatorizado por bloques, pero al recopilar las observaciones durante seis días no ha sido posible aplicar todos los tratamientos en cada día, sino que sólo se han podido aplicar dos de las cuatro soluciones cada día. Se decide utilizar un diseño en bloques incompletos balanceado, donde  $I = 4$  y  $K = 2$ .

Un posible diseño para estos parámetros lo proporciona la tabla correspondiente al Diseño 5 del Fichero-Adjunto, con  $R = 3$ ,  $J = 6$  y  $\lambda = 1$ . La disposición del diseño y las observaciones obtenidas se muestran en la siguiente tabla.

El objetivo principal es estudiar la efectividad en el retraso del crecimiento de bacterias utilizando cuatro soluciones, por lo que se trata de un factor con cuatro niveles. Sin embargo, como los días son una fuente de variabilidad potencial, consideramos un factor bloque con seis niveles.

- Variable respuesta: Número de bacterias

	Días					
Soluciones	1	2	3	4	5	6
Solución 1	12	24	31			
Solución 2	21				20	21
Solución 3			19	18		19
Solución 4		15		19	47	

Tabla 8; Datos del Ejercicio Propuesto 4

- Factor: Soluciones que tiene cuatro niveles. Es un factor de efectos fijos ya que viene decidido qué niveles concretos se van a utilizar.
- Bloque: Días que tiene seis niveles. Es un factor de efectos fijos ya que viene decidido qué niveles concretos se van a utilizar.
- Modelo incompleto: Todos los tratamientos no se prueban en cada bloque.
- Tamaño del experimento: Número total de observaciones (12).

Podemos introducir los datos directamente en R de forma manual o introducirlos previamente en un archivo de texto o Excel y leerlos en R.

```
bacterias<-read.table("bacterias.txt",header = TRUE);bacterias
```

```

      y Soluciones Dias
1  12             1    1
2  24             1    2
3  31             1    3
4  21             2    1
5  20             2    5
6  21             2    6
7  19             3    3
8  18             3    4
9  19             3    6
10 15             4    2
11 19             4    4
12 47             4    5

```

A continuación debemos transformar tanto la columna de los tratamientos como la de los bloques en un factor para poder realizar los cálculos posteriores adecuadamente.

```

bacterias$Soluciones=factor(bacterias$Soluciones)
bacterias$Dias=factor(bacterias$Dias)

```

Para poder analizar los datos mediante un diseño BIB debemos instalar y cargar dos paquetes de R especializados en este tipo de diseños:

```
library(daewr)
library(AlgDesign)
```

La función “**BIBsize(t , k)**” de la librería daewr nos permite saber si el diseño puede realizarse. Calcula los parámetros del diseño donde

- t = número de niveles del factor tratamiento.
- k = número de tratamientos por bloque.

Ejecutamos:

```
BIBsize(t=4,k=2)
```

Posible BIB design with b= 6 and r= 3 lambda= 1

El análisis de este modelo lo podemos realizar en R de dos formas:

1. Realizaremos el análisis evaluando primero el efecto de los tratamientos y después el de los bloques utilizando dos funciones
- Para evaluar el efecto de los tratamientos, la suma de cuadrados de tratamientos debe ajustarse por bloques, por lo tanto primero se introducen los bloques y después los tratamientos.
  - Para calcular la tabla ANOVA hacemos uso de la función “**aov**” (**aov(y ~ A + B, data=mydataframe)** asume suma de cuadrados tipo I) de la siguiente forma:

donde:

- y = nombre de la columna de las observaciones
- Soluciones = nombre de la columna en la que están representados los tratamientos
- Dias = nombre de la columna en la que están representados los bloques
- data = data.frame en el que están guardados los datos

```
mod1<-aov(y~Dias+Soluciones,data=bacterias);mod1
```

Call:

```
aov(formula = y ~ Dias + Soluciones, data = bacterias)
```

Terms:

	Dias	Soluciones	Residuals
Sum of Squares	387.6667	123.2500	396.7500
Deg. of Freedom	5	3	3

Residual standard error: 11.5  
Estimated effects may be unbalanced

```
# resumen con la función summary
summary(mod1)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Dias	5	387.7	77.53	0.586	0.720
Soluciones	3	123.3	41.08	0.311	0.819
Residuals	3	396.7	132.25		

El valor del estadístico de contraste de igualdad de Soluciones,  $F = 0.311$ , deja a su derecha un p-valor 0.819, mayor que el nivel de significación del 5%, por lo que no se rechaza la Hipótesis Nula de igualdad de tratamientos. Por lo tanto el tipo de solución para lavar los envases de la leche no influye en el retraso del crecimiento de bacterias.

- Para evaluar el efecto de los bloques, la suma de cuadrados de bloques debe ajustarse por los tratamientos, por lo tanto primero se introducen los tratamientos y después los bloques:

```
mod2<-aov(y ~ Soluciones + Dias, data = bacterias );mod2
```

Call:

```
aov(formula = y ~ Soluciones + Dias, data = bacterias)
```

Terms:

	Soluciones	Dias	Residuals
Sum of Squares	113.6667	397.2500	396.7500
Deg. of Freedom	3	5	3

Residual standard error: 11.5  
Estimated effects may be unbalanced

```
summary(mod2)
```



	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Soluciones	3	113.7	37.89	0.286	0.834
Dias	5	397.2	79.45	0.601	0.712
Residuals	3	396.7	132.25		

El valor del estadístico de contraste de igualdad de **Días**, **F = 0.601**, deja a su derecha un p-valor 0.712, mayor que el nivel de significación del 5%, por lo que no se rechaza la Hipótesis nula de igualdad de bloques. Por lo tanto los días en los que se realiza la prueba para lavar los envases de la leche no influyen en el retraso del crecimiento de bacterias.

2. Realizaremos el análisis evaluando tanto para los tratamientos como para los bloques ejecutando solo una función.

```
library(car)
```

Loading required package: carData

```
mod3 <- lm(y ~ Soluciones + Dias, data = bacterias);mod3
```

Call:

```
lm(formula = y ~ Soluciones + Dias, data = bacterias)
```

Coefficients:

(Intercept)	Soluciones2	Soluciones3	Soluciones4	Dias2	Dias3
20.000	-7.000	-6.750	1.750	-1.375	8.375
Dias4	Dias5	Dias6			
1.000	16.125	6.875			

```
car::Anova(mod3,type="III")
```

Anova Table (Type III tests)

Response: y

	Sum Sq	Df	F value	Pr(>F)
(Intercept)	533.33	1	4.0328	0.1382
Soluciones	123.25	3	0.3106	0.8187
Dias	397.25	5	0.6008	0.7118
Residuals	396.75	3		

Los resultados obtenidos coinciden con los realizados primero a los tratamientos y después a los bloques.

## 7 Diseño en Cuadrados Latinos

### Supuesto práctico 5

Se estudia el rendimiento de un proceso químico en seis tiempos de reposo, A, B, C, D, E y F. Para ello, se consideran seis lotes de materia prima que reaccionan con seis concentraciones de ácido distintas, de manera que cada lote de materia prima en cada concentración de ácido se somete a un tiempo de reposo. Tanto la asignación de los tiempos de reposo a los lotes de materia prima, como la concentración de ácido, se hizo de forma aleatoria. Los datos del rendimiento del proceso químico se muestran en la siguiente tabla.

Concentraciones de ácido						
Lote	1	2	3	4	5	6 7
Lote 1	12 A	24 B	10 C	18 D	21 E	18 F
Lote 2	21 B	26 C	24 D	16 E	20 F	21 A
Lote 3	20 C	16 D	19 E	18 F	16 A	19 B
Lote 4	22 D	15 E	14 F	19 A	27 B	17 C
Lote 5	15 E	13 F	17 A	25 B	21 C	22 D
Lote 6	17 F	11 A	12 B	22 C	14 D	20 E

```
latino<-read.table("Lotes2.txt",header = TRUE,dec = ",")
latino
```

	Observaciones	Lote	Concentraciones	Tiempo_de_reposo
1		12 Lote1	1	A
2		24 Lote1	2	B
3		10 Lote1	3	C
4		18 Lote1	4	D
5		21 Lote1	5	E
6		18 Lote1	6	F
7		21 Lote2	1	B
8		26 Lote2	2	C
9		24 Lote2	3	D
10		16 Lote2	4	E
11		20 Lote2	5	F
12		21 Lote2	6	A
13		20 Lote3	1	C
14		16 Lote3	2	D
15		19 Lote3	3	E
16		18 Lote3	4	F
17		16 Lote3	5	A

18	19 Lote3	6	B
19	22 Lote4	1	D
20	15 Lote4	2	E
21	14 Lote4	3	F
22	19 Lote4	4	A
23	27 Lote4	5	B
24	17 Lote4	6	C
25	15 Lote5	1	E
26	13 Lote5	2	F
27	17 Lote5	3	A
28	25 Lote5	4	B
29	21 Lote5	5	C
30	22 Lote5	6	D
31	17 Lote6	1	F
32	11 Lote6	2	A
33	12 Lote6	3	B
34	22 Lote6	4	C
35	14 Lote6	5	D
36	20 Lote6	6	E

A continuación debemos transformar tanto la columna de los tratamientos como la de los bloques en un factor para poder realizar los cálculos posteriores adecuadamente.

```
latino$Lote<-factor(latino$Lote)
latino$Concentraciones<-factor(latino$Concentraciones)
latino$Tiempo_de_reposo<-factor(latino$Tiempo_de_reposo)
```

Para calcular la tabla ANOVA primero hacemos uso de la función “aov” de la siguiente forma:

```
mod1<-aov(Observaciones~Lote+Concentraciones+Tiempo_de_reposo,data = latino)
mod1
```

Call:

```
aov(formula = Observaciones ~ Lote + Concentraciones + Tiempo_de_reposo,
     data = latino)
```

Terms:

	Lote	Concentraciones	Tiempo_de_reposo	Residuals
Sum of Squares	99.5556	70.5556	117.8889	346.5556
Deg. of Freedom	5	5	5	20

Residual standard error: 4.162665

Estimated effects may be unbalanced

```
summary(mod1)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Lote	5	99.6	19.91	1.149	0.368
Concentraciones	5	70.6	14.11	0.814	0.553
Tiempo_de_reposo	5	117.9	23.58	1.361	0.281
Residuals	20	346.6	17.33		

Observando los valores de los p-valores, 0.281, 0.368 y 0.553; mayores respectivamente que el nivel de significación del 5%, deducimos que ningún efecto es significativo.

## 8 Diseño en Cuadrados Greco-Latinos

### Supuesto práctico 6

Para comprobar el rendimiento de un proceso químico en cinco tiempos de reposo, se consideran cinco lotes de materia prima que reaccionan con cinco concentraciones de ácido distintas a cinco temperaturas distintas, de manera que cada lote de materia prima con cada concentración de ácido y cada temperatura se someten a un tiempo de reposo. Tanto la asignación de los tiempos de reposo a los lotes de materia prima, como las concentraciones de ácido, y las temperaturas, se hizo de forma aleatoria. En este estudio el científico considera que tanto los lotes de materia prima, las concentraciones y las temperaturas pueden influir en el rendimiento del proceso, por lo que los considera como variables de bloque cada una con cinco niveles y decide plantear un diseño por cuadrados greco-latinos como el que muestra en la siguiente tabla.

Rendimiento					
	Concentraciones de ácido				
lote	1	2	3	4	5
lote 1	26 A $\alpha$	21 B $\beta$	19 C $\gamma$	13 D $\delta$	21 E $\eta$
lote 2	22 B $\gamma$	26 C $\delta$	24 D $\eta$	16 E $\alpha$	20 A $\beta$
lote 3	29 C $\eta$	26 D $\alpha$	19 E $\beta$	18 A $\gamma$	16 B $\delta$
lote 4	32 D $\beta$	15 E $\gamma$	14 A $\delta$	19 B $\eta$	27 C $\alpha$
lote 5	25 E $\delta$	18 A $\eta$	19 B $\alpha$	25 C $\beta$	21 D $\gamma$

Tabla de datos del Supuesto

Práctico 6

Tenemos en cuenta que para que el ejercicio esté realizado de forma correcta los datos tienen que estar introducidos tal y como vienen en la imagen, es decir, las observaciones en una sola columna y a continuación especificado su tratamiento, su bloque correspondiente y después la letra latina y griega correspondiente (En este caso hemos cambiado las letras griegas como las últimas del alfabeto latino por facilidad a la hora de escribirlas).

```
greco <- read.table("Lotes_2.txt", header = TRUE, dec= ",")
greco
```

	Observaciones	Lotes	Concentraciones	Tiempo_de_reposo	Temperaturas
1	26	Lote1	1	A	Z
2	21	Lote1	2	B	Y
3	19	Lote1	3	C	X
4	13	Lote1	5	D	W
5	21	Lote1	5	E	V
6	22	Lote2	1	B	X
7	26	Lote2	2	C	W
8	24	Lote2	3	D	V
9	16	Lote2	4	E	Z
10	20	Lote2	5	A	Y
11	29	Lote3	1	C	V
12	26	Lote3	2	D	Z
13	19	Lote3	3	E	Y
14	18	Lote3	4	A	X
15	16	Lote3	5	B	W
16	32	Lote4	1	D	Y
17	15	Lote4	2	E	X
18	14	Lote4	3	A	W
19	19	Lote4	4	B	V
20	27	Lote4	5	C	Z
21	25	Lote5	1	E	W
22	18	Lote5	2	A	V
23	19	Lote5	3	B	Z
24	25	Lote5	4	C	Y
25	21	Lote5	5	D	X

A continuación debemos transformar tanto la columna de los tratamientos como la de los bloques en un factor para poder realizar los cálculos posteriores adecuadamente.

Para calcular la tabla ANOVA primero hacemos uso de la función “aov” de la siguiente forma:

```
greco$Lote <- factor(greco$Lote)
greco$Temperaturas <- factor(greco$Temperaturas)
greco$Tiempo_de_reposo <- factor(greco$Tiempo_de_reposo)
greco$Concentraciones <- factor(greco$Concentraciones)
mod1 <- aov(Observaciones~ Lote + Concentraciones + Tiempo_de_reposo + Temperaturas, data = greco)
mod1
```

Call:

```
aov(formula = Observaciones ~ Lote + Concentraciones + Tiempo_de_reposo +
    Temperaturas, data = greco)
```

Terms:

	Lote	Concentraciones	Tiempo_de_reposo	Temperaturas
Sum of Squares	9.7600	207.7607	155.0085	97.2516
Deg. of Freedom	4	4	4	4
	Residuals			
Sum of Squares	100.7792			
Deg. of Freedom	8			

Residual standard error: 3.549281

Estimated effects may be unbalanced

donde:

- Observaciones: Nombre de la columna de las observaciones
  - Lote: Nombre de la columna en la que están representados los tratamientos
  - Concentraciones = Nombre de la columna en la que está representado el primer factor bloque
  - Tiempo\_de\_reposo = Nombre de la columna en la que está representado el segundo factor bloque (letras latinas)
  - Temperaturas: Nombre de la columna en la que está representado el tercer factor bloque
- Data: data.frame en el que están guardados los datos

y posteriormente mostramos un resumen de los resultados con la función “summary” (verdadera tabla ANOVA):

```
summary(mod1)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Lote	4	9.76	2.44	0.194	0.9349
Concentraciones	4	207.76	51.94	4.123	0.0420 *

```
Tiempo_de_reposo  4 155.01   38.75   3.076 0.0825 .
Temperaturas      4  97.25   24.31   1.930 0.1988
Residuals         8 100.78   12.60
```

---

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Observando los valores de los p-valores, 0.1988, 0.0825, 0.0420 y 0.9349, deducimos que el único efecto significativo, al nivel de significación del 5%, es el efecto de las distintas concentraciones sobre el rendimiento del proceso químico.

## 9 Diseño en Cuadrados de Youden

### Supuesto práctico 7

Consideremos de nuevo el experimento sobre el rendimiento de un proceso químico en el que se está interesado en estudiar seis tiempos de reposo, A, B, C, D, E y F y se desea eliminar estadísticamente el efecto de los lotes materia prima y de las concentraciones de ácido distintas. Pero supongamos que sólo se dispone de cinco tipos de concentraciones. Para analizar este experimento se decidió utilizar un cuadrado de Youden con seis filas (los lotes de materia prima), cinco columnas (las distintas concentraciones) y seis letras latinas (los tiempos de reposo). Los datos correspondientes se muestran en la siguiente tabla.

Tenemos en cuenta que para que el ejercicio esté realizado de forma correcta los datos tienen que estar introducidos tal y como vienen en la imagen, es decir, las observaciones en una sola columna y a continuación especificado su tratamiento, su bloque correspondiente y después la letra latina correspondiente.

```
youden <- read.table("Lotes3.txt", header = TRUE)
youden
```

	Observaciones	Lote	Concentraciones	Tiempo_de_reposo
1	12	Lote1	1	A
2	24	Lote1	2	B
3	10	Lote1	3	C
4	18	Lote1	4	D
5	21	Lote1	5	E
6	21	Lote2	1	B
7	26	Lote2	2	C
8	24	Lote2	3	D
9	16	Lote2	4	E
10	20	Lote2	5	F

	<b>Concentraciones de ácido</b>				
<b>Lote</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>
Lote 1	12 A	24 B	10 C	18 D	21 E
Lote 2	21 B	26 C	24 D	16 E	20 F
Lote 3	20 C	16 D	19 E	18 F	16 A
Lote 4	22 D	15 E	14 F	19 A	27 B
Lote 5	15 E	13 F	17 A	25 B	21 C
Lote 6	17 F	11 A	12 B	22 C	14 D

Tabla 8; Tabla de  
datos del Supuesto Práctico 7.

11	20 Lote3	1	C
12	16 Lote3	2	D
13	19 Lote3	3	E
14	18 Lote3	4	F
15	16 Lote3	5	A
16	22 Lote4	1	D
17	15 Lote4	2	E
18	14 Lote4	3	F
19	19 Lote4	4	A
20	27 Lote4	5	B
21	15 Lote5	1	E
22	13 Lote5	2	F
23	17 Lote5	3	A
24	25 Lote5	4	B
25	21 Lote5	5	C
26	17 Lote6	1	F
27	11 Lote6	2	A
28	12 Lote6	3	B
29	22 Lote6	4	C
30	14 Lote6	5	D



A continuación debemos transformar tanto la columna de los tratamiento como la de los bloques en un factor para podemos realizar los cálculos posteriores adecuadamente.

```
youden$Lote <- factor(youden$Lote)
youden$Concentraciones <- factor(youden$Concentraciones)
youden$Tiempo_de_reposo <- factor(youden$Tiempo_de_reposo)
```

Para cada factor realizamos una tabla ANOVA:

- **Factor principal:**

Para evaluar el efecto de los tratamientos, la suma de cuadrados de tratamientos debe ajustarse por bloques, por lo tanto primero se introducen los bloques y después los tratamientos.

Para calcular la tabla ANOVA hacemos uso de la función “aov” (asume suma de cuadrados tipo I) de la siguiente forma:

```
mod1 <- aov(Observaciones~ Tiempo_de_reposo + Lote + Concentraciones, data = youden )
mod1
```

Call:

```
aov(formula = Observaciones ~ Tiempo_de_reposo + Lote + Concentraciones,
     data = youden)
```

Terms:

	Tiempo_de_reposo	Lote	Concentraciones	Residuals
Sum of Squares	151.76667	112.73333	61.66667	282.00000
Deg. of Freedom	5	5	4	15

Residual standard error: 4.335897

Estimated effects may be unbalanced

```
summary(mod1)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Tiempo_de_reposo	5	151.77	30.35	1.615	0.216
Lote	5	112.73	22.55	1.199	0.356
Concentraciones	4	61.67	15.42	0.820	0.532
Residuals	15	282.00	18.80		

El p-valor, **0.532**, es mayor que el nivel de significación del 5%, deducimos que el **factor principal: Concentraciones no es significativo**.

- **Factor Bloque: Lotes.** Para evaluar el efecto del primero de los bloques, la suma de cuadrados de bloques debe ajustarse por los tratamientos, por lo tanto primero se introducen los tratamientos y después los bloques:

```
mod2 <- aov(Observaciones~ Concentraciones +Tiempo_de_reposo + Lote , data = youden )
mod2
```

Call:

```
aov(formula = Observaciones ~ Concentraciones + Tiempo_de_reposo +
     Lote, data = youden)
```

Terms:

	Concentraciones	Tiempo_de_reposo	Lote	Residuals
Sum of Squares	61.66667	151.76667	112.73333	282.00000
Deg. of Freedom	4	5	5	15

Residual standard error: 4.335897

Estimated effects may be unbalanced

```
summary(mod2)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Concentraciones	4	61.67	15.42	0.820	0.532
Tiempo_de_reposo	5	151.77	30.35	1.615	0.216
Lote	5	112.73	22.55	1.199	0.356
Residuals	15	282.00	18.80		

El p-valor, **0.356**, es mayor que el nivel de significación del 5%, deducimos que el **Factor Bloque: Lotes no es significativo.**

- **Factor Bloque:Tiempo\_de\_reposo** Para evaluar el efecto del segundo bloque, la suma de cuadrados de bloques debe ajustarse también por los tratamientos, por lo tanto primero se introducen los tratamientos y después los bloques:

```
mod3 <- aov(Observaciones~ Concentraciones + Lote +Tiempo_de_reposo , data = youden )
mod3
```

Call:

```
aov(formula = Observaciones ~ Concentraciones + Lote + Tiempo_de_reposo,
     data = youden)
```

Terms:

	Concentraciones	Lote	Tiempo_de_reposo	Residuals
Sum of Squares	61.66667	111.36667	153.13333	282.00000
Deg. of Freedom	4	5	5	15

Residual standard error: 4.335897

Estimated effects may be unbalanced

```
summary(mod3)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Concentraciones	4	61.67	15.42	0.820	0.532
Lote	5	111.37	22.27	1.185	0.362
Tiempo_de_reposo	5	153.13	30.63	1.629	0.213
Residuals	15	282.00	18.80		

El p-valor es **0.213**; mayor que el nivel de significación del 5%, deducimos que Factor **Bloque:Tiempo\_de\_reposo** no es significativo.

## 10 Diseñp Factoriales

### 10.1 Diseños factoriales con dos factores

En primer lugar vamos a estudiar los diseños más simples, es decir aquellos en los que intervienen sólo dos factores. Supongamos que hay  $a$  niveles para el factor A y  $b$  niveles del factor B, cada réplica del experimento contiene todas las posibles combinaciones de tratamientos, es decir contiene los  $ab$  tratamientos posibles.

#### 10.1.1 El modelo sin replicación

El modelo estadístico para este diseño es:

$$y_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + u_{ij}, \quad i = 1, \dots, a; \quad j = 1, \dots, b$$

Expresión 16: Modelo estadísticodel diseño factorial de dos factores sin replicación

Donde:

- $y_{ij}$ : Representa la observación correspondiente al nivel (i) del factor A y al nivel (j) del factor B.
- $\mu$ : Efecto constante, común a todos los niveles de los factores, denominado media global.
- $\tau_i$ : Efecto producido por el nivel i-ésimo del factor A, ( $\sum_i \tau_i = 0$ ).
- $\beta_j$ : Efecto producido por el nivel j-ésimo del factor B, ( $\sum_j \beta_j = 0$ ).
- $(\tau\beta)_{ij}$ : Efecto producido por la interacción entre  $A \times B$ ,  $\sum_i (\tau\beta)_{ij} = \sum_j (\tau\beta)_{ij} = 0$ .
- $u_{ij}$  son vv aa. independientes con distribución  $N(0, \sigma)$ .

### Supuesto práctico 8

En unos laboratorios se está investigando sobre el tiempo de supervivencia de unos animales a los que se les suministra al azar tres tipos de venenos y cuatro antídotos distintos. Se pretende estudiar si los tiempos de supervivencia de los animales varían en función de las combinaciones veneno-antídoto. Los datos que se recogen en la tabla adjunta son los tiempos de supervivencia en horas.

	Antídoto			
Veneno	Antídoto 1	Antídoto 2	Antídoto 3	Antídoto 4
Veneno 1	4.5	11.0	4.5	7.1
Veneno 2	2.9	6.1	3.5	10.2
Veneno 3	2.1	3.7	2.5	3.6

Tabla 17; Tabla de datos del Supuesto Práctico 8.

El objetivo principal es estudiar la influencia de tres tipos de venenos y 4 tipos de antídotos en el tiempo de supervivencia de unos determinados animales, por lo que se trata de un modelo con dos factores: el veneno (con tres niveles) y el antídoto (con cuatro niveles). La variable que va a medir las diferencias entre los tratamientos es el tiempo que sobreviven los animales. Se combinan todos los niveles de los dos factores por lo que tenemos en total doce tratamientos.

Tenemos en cuenta que para que el ejercicio esté realizado de forma correcta los datos tienen que estar introducidos tal y como vienen en la imagen, es decir, las observaciones en una sola columna y a continuación especificado sus factores correspondientes.

```
factorial <- read.table("laboratorio.txt", header = TRUE)
factorial
```

```

      Tiempo Veneno Antidoto
1         4.5      1        1
2         2.9      2        1
```

3	2.1	3	1
4	11.0	1	2
5	6.1	2	2
6	3.7	3	2
7	4.5	1	3
8	3.5	2	3
9	2.5	3	3
10	7.1	1	4
11	10.2	2	4
12	3.6	3	4

A continuación debemos transformar todas las columnas que contienen a los factores en un factor para podemos realizar los cálculos posteriores adecuadamente.

```
factorial$Antidoto <- factor(factorial$Antidoto)
factorial$Veneno <- factor(factorial$Veneno)
```

Para calcular la tabla ANOVA primero hacemos uso de la función “aov” de la siguiente forma

```
mod <- aov(Tiempo~ Veneno + Antidoto , data = factorial )
mod
```

Call:

```
aov(formula = Tiempo ~ Veneno + Antidoto, data = factorial)
```

Terms:

	Veneno	Antidoto	Residuals
Sum of Squares	30.58667	39.40917	23.89333
Deg. of Freedom	2	3	6

Residual standard error: 1.995551

Estimated effects may be unbalanced

```
summary(mod)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Veneno	2	30.59	15.293	3.840	0.0844 .
Antidoto	3	39.41	13.136	3.299	0.0995 .
Residuals	6	23.89	3.982		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Esta Tabla ANOVA recoge la descomposición de la varianza considerando como fuente de variación los doce tratamientos o grupos que se forman al combinar los niveles de los dos factores. Mediante esta tabla se puede estudiar si varían los tiempos que sobreviven los animales en función de las combinaciones veneno-antídoto. Es decir, se pueden estudiar si existen diferencias significativas entre los tiempos medios de supervivencia con los distintos tipos de venenos y antídotos, pero no se puede estudiar si la efectividad de los antídotos es la misma para todos los venenos. Observando los p-valores, **0.084 y 0.099**; mayores respectivamente que el nivel de significación del 5%, deducimos que ningún efecto es significativo. Por lo tanto, no existen diferencias en los tiempos medios de supervivencia de los animales, en función de la pareja veneno-antídoto que se les suministra.

### 10.1.2 El modelo con replicación

El modelo estadístico para este diseño es:

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + u_{ijk}, \quad i = 1, \dots, a; \quad j = 1, \dots, b; \quad k = 1, \dots, r$$

Expresión 17: Modelo estadístico del diseño factorial de dos factores con replicación

### Supuesto práctico 9

Consideremos el supuesto práctico anterior en el que realizamos dos réplicas por cada tratamiento. Los datos que se recogen en la tabla adjunta son los tiempos de supervivencia en horas de unos animales a los que se les suministra al azar tres venenos y cuatro antídotos. El objetivo es estudiar qué antídoto es el adecuado para cada veneno.

	<b>Antídoto</b>			
<b>Veneno</b>	<b>Antídoto 1</b>	<b>Antídoto 2</b>	<b>Antídoto 3</b>	<b>Antídoto 4</b>
<b>Veneno 1</b>	4.5	11.0	4.5	7.1
	4.3	7.2	7.6	6.2
<b>Veneno 2</b>	2.9	6.1	3.5	10.2
	2.3	12.4	4.0	3.8
<b>Veneno 3</b>	2.1	3.7	2.5	3.6
	2.3	2.9	2.2	3.3

Tabla 19; Tabla de datos del Supuesto Práctico 9.

enemos en cuenta que para que el ejercicio esté realizado de forma correcta los datos tienen que estar introducidos tal y como vienen en la imagen, es decir, las observaciones en una sola columna y a continuación especificado sus factores correspondientes.

```
factorial <- read.table("laboratorio2.txt", header = TRUE)
factorial
```

	Tiempo	Veneno	Antidoto
1	4.5	1	1
2	4.3	1	1
3	2.9	2	1
4	2.3	2	1
5	2.1	3	1
6	2.3	3	1
7	11.0	1	2
8	7.2	1	2
9	6.1	2	2
10	12.4	2	2
11	3.7	3	2
12	2.9	3	2
13	4.5	1	3
14	7.6	1	3
15	3.5	2	3
16	4.0	2	3
17	2.5	3	3
18	2.2	3	3
19	7.1	1	4
20	6.2	1	4
21	10.2	2	4
22	3.8	2	4
23	3.6	3	4
24	3.3	3	4

A continuación debemos transformar todas las columnas que contienen a los factores en un factor para podemos realizar los cálculos posteriores adecuadamente.

```
factorial$Veneno <- factor(factorial$Veneno)
factorial$Antidoto <- factor(factorial$Antidoto)
```

Para calcular la tabla ANOVA primero hacemos uso de la función “aov” de la siguiente forma:

```
mod <- aov(Tiempo~ Veneno * Antidoto , data = factorial )
mod
```

Call:

```
aov(formula = Tiempo ~ Veneno * Antidoto, data = factorial)
```

Terms:

	Veneno	Antidoto	Veneno:Antidoto	Residuals
Sum of Squares	60.44333	60.26167	20.36333	53.51000
Deg. of Freedom	2	3	6	12

Residual standard error: 2.111674

Estimated effects may be unbalanced

```
summary(mod)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Veneno	2	60.44	30.222	6.777	0.0107 *
Antidoto	3	60.26	20.087	4.505	0.0245 *
Veneno:Antidoto	6	20.36	3.394	0.761	0.6138
Residuals	12	53.51	4.459		

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

La Tabla ANOVA muestra las filas de Tipo\_veneno, Tipo\_antídoto y Tipo\_veneno\*Tipo\_antídoto que corresponde a la variabilidad debida a los efectos de cada uno de los factores y de la interacción entre ambos.

Las preguntas que nos planteamos son: ¿Los venenos son igual de peligrosos? ¿Y los antídotos son igual de efectivos? La efectividad de los antídotos, ¿es la misma para todos los venenos? Para responder a estas preguntas, comenzamos comprobando si el efecto de los antídotos es el mismo para todos los venenos. Para ello observamos el valor del estadístico ( $F_{exp} = 0.761$ ) que contrasta la hipótesis correspondiente a la interacción entre ambos factores ( $H_0 : (\tau\beta)_{ij} = 0$ ). Dicho valor deja a la derecha un Sig. = 0.614, mayor que el nivel de significación 0.05. Por lo tanto la interacción entre ambos factores no es significativa y debemos eliminarla del modelo. Construimos de nuevo la Tabla ANOVA en la que sólo figurarán los efectos principales

```
mod <- aov(Tiempo~ Veneno + Antidoto , data = factorial )
mod
```



Call:

```
aov(formula = Tiempo ~ Veneno + Antidoto, data = factorial)
```

Terms:

	Veneno	Antidoto	Residuals
Sum of Squares	60.44333	60.26167	73.87333
Deg. of Freedom	2	3	18

Residual standard error: 2.025851

Estimated effects may be unbalanced

```
summary(mod)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Veneno	2	60.44	30.222	7.364	0.0046 **
Antidoto	3	60.26	20.087	4.894	0.0117 *
Residuals	18	73.87	4.104		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Esta tabla muestra dos únicas fuentes de variación, los efectos principales de los dos factores (Tipo\_veneno y Tipo\_antídoto), y se ha suprimido la interacción entre ambos. Se observa que el valor de la Suma de Cuadrados del error de este modelo (73.873) se ha formado con los valores de las Sumas de cuadrados del error y de la interacción del modelo anterior ( $20.363 + 53.510 = 73.873$ ). Observando los valores de los p-valores, 0.0046 y 0.0117 asociados a los contrastes principales, se deduce que los dos efectos son significativos a un nivel de significación del 5%. Deducimos que ni la gravedad de los venenos es la misma, ni la efectividad de los antídotos, pero dicha efectividad no depende del tipo de veneno con el que se administre ya que la interacción no es significativa.

## 10.2 Diseños factoriales con tres factores

### 10.2.1 El modelo sin replicación

#### Supuesto práctico 10

En una fábrica de refrescos está haciendo unos estudios en la planta embotelladora. El objetivo es obtener más uniformidad en el llenado de las botellas. La máquina de llenado teóricamente llena cada botella a la altura correcta, pero en la práctica hay variación, y la embotelladora desea entender mejor las fuentes de esta variabilidad para eventualmente reducirla. En el proceso se pueden controlar tres factores durante el proceso de llenado: El % de carbonato

(factor A), la presión del llenado (factor B) y el número de botellas llenadas por minuto que llamaremos velocidad de la línea (factor C). Se consideran tres niveles para el factor A (10%, 12%, 14%), dos niveles para el factor B (25psi, 30psi) y dos niveles para el factor C (200bpm, 250bpm). Los datos recogidos de la desviación de la altura objetivo se muestran en la tabla adjunta

	Presión (B)			
	25 psi		30 psi	
	Velocidad (C)		Velocidad (C)	
% de Carbono (A)	200	250	200	250
10	10	3	5	-1
12	11	2	5	-3
14	2	4	-3	1

Tabla 22; Tabla

de datos del Supuesto Práctico 10.

Tenemos en cuenta que para que el ejercicio esté realizado de forma correcta los datos tienen que estar introducidos tal y como vienen en la imagen, es decir, las observaciones en una sola columna y a continuación especificado sus factores correspondientes.

```
factorial <- read.table("embotelladora.txt", header = TRUE)
factorial
```

	Altura	Carbono	Presion	Velocidad
1	10	10	25	200
2	11	12	25	200
3	2	14	25	200
4	3	10	25	250
5	2	12	25	250
6	4	14	25	250
7	5	10	30	200
8	5	12	30	200
9	-3	14	30	200
10	-1	10	30	250
11	-3	12	30	250
12	1	14	30	250

A continuación debemos transformar la tres columnas en factores para poder realizar los

cálculos posteriores adecuadamente.

```
factorial$Carbono <- factor(factorial$Carbono)
factorial$Velocidad <- factor(factorial$Velocidad)
factorial$Presion <- factor(factorial$Presion)
```

Para calcular la tabla ANOVA primero hacemos uso de la función “aov” de la siguiente forma:

```
mod <- aov(Altura~ Carbono + Presion + Velocidad + Carbono*Presion +
          Carbono*Velocidad + Presion*Velocidad , data = factorial )
mod
```

Call:

```
aov(formula = Altura ~ Carbono + Presion + Velocidad + Carbono *
     Presion + Carbono * Velocidad + Presion * Velocidad, data = factorial)
```

Terms:

	Carbono	Presion	Velocidad	Carbono:Presion	Carbono:Velocidad
Sum of Squares	24.50000	65.33333	48.00000	1.16667	75.50000
Deg. of Freedom	2	1	1	2	2

	Presion:Velocidad	Residuals
Sum of Squares	1.33333	0.16667
Deg. of Freedom	1	2

Residual standard error: 0.2886751  
Estimated effects may be unbalanced

donde:

- **Altura:** Nombre de la columna de las observaciones
- **Carbono:** Nombre de la columna en la que está representado el primer factor
- **Presion:** Nombre de la columna en la que está representado el segundo factor
- **Velocidad:** Nombre de la columna en la que está representado el tercer factor
- **CarbonoPresion, Carbono× Velocidad y Presion Velocidad** hace referencia a las distintas interacciones.

data= data.frame en el que están guardados los datos

```
summary(mod)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Carbono	2	24.50	12.25	147	0.00676 **

Presion	1	65.33	65.33	784	0.00127	**
Velocidad	1	48.00	48.00	576	0.00173	**
Carbono:Presion	2	1.17	0.58	7	0.12500	
Carbono:Velocidad	2	75.50	37.75	453	0.00220	**
Presion:Velocidad	1	1.33	1.33	16	0.05719	.
Residuals	2	0.17	0.08			

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

La Tabla ANOVA muestra las filas de Carbono, Presión, Velocidad, Carbono\* Presión, Carbono\* Velocidad y Presión\* Velocidad que corresponden a la variabilidad debida a los efectos de cada uno de los factores y a las interacciones de orden dos entre ambos. En dicha Tabla se indica que para un nivel de significación del 5% los efectos que no son significativos del modelo planteado son las interacciones entre los factores Carbono\* Presión y Presión\*Velocidad ya que los p-valores correspondientes a estos efectos son 0.125 y 0.057 mayores que el nivel de significación.

donde los efectos deben cumplir las condiciones expuestas anteriormente. Para resolverlo suprimimos la interacción Carbono\* Presión. La tabla ANOVA que corresponde a este modelo es la siguiente:

```
mod <- aov(Altura~ Carbono + Presion + Velocidad + Carbono*Velocidad +
           Presion*Velocidad , data = factorial )
summary(mod)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Carbono	2	24.50	12.25	36.75	0.002664	**
Presion	1	65.33	65.33	196.00	0.000151	***
Velocidad	1	48.00	48.00	144.00	0.000276	***
Carbono:Velocidad	2	75.50	37.75	113.25	0.000301	***
Presion:Velocidad	1	1.33	1.33	4.00	0.116117	
Residuals	4	1.33	0.33			

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Todos los efectos de este último modelo planteado son significativos y por lo tanto es en este modelo donde vamos a realizar el estudio. Existen diferencias significativas entre los distintos porcentajes del Carbono, los dos tipos de presión, las dos velocidades de llenado y la interacción entre el porcentaje de Carbono y la Velocidad de llenado.

### 10.3 Diseño factorial de tres factores con replicación

#### Supuesto práctico 11

Consideremos el supuesto práctico anterior en el que realizamos dos réplicas por cada tratamiento. En la Tabla adjunta se muestran los datos recogidos de la desviación de la altura objetivo de las botellas de refresco. En el proceso de llenado, la embotelladora puede controlar tres factores durante el proceso: El porcentaje de carbonato (factor A) con tres niveles (10%, 12%, 14%), la presión del llenado (factor B) con dos niveles (25psi, 30psi) y el número de botellas llenadas por minuto que llamaremos velocidad de la línea (factor C) con dos niveles (200bpm, 250bpm).

	Presión (B)			
	25 psi		30 psi	
	Velocidad (C)		Velocidad (C)	
% de Carbono (A)	200	250	200	250
10	10	3	5	-1
	20	5	9	-3
12	11	2	5	-3
	9	5	4	2
14	2	4	-3	1
	-1	7	-2	3

Tabla 24; Tabla de datos del Supuesto Práctico 11.

La variable respuesta de este experimento es la Desviación que se produce de la altura objetivo en el llenado en las botellas de refresco. Los factores son: Porcentaje de Carbono que presenta tres niveles 10%, 12% y 14%; Presión, con dos niveles 25 psi y 30 psi y Velocidad, con dos niveles 200 y 250. Los niveles de los factores han sido fijados por el experimentador, por lo que todos los factores son de efectos fijos. Se trata de un diseño trifactorial de efectos fijos, donde el número de tratamientos es  $3 \times 2 \times 2 = 12$  y el número de observaciones 24.

```
factorial <- read.table("embotelladora2.txt", header = TRUE)
factorial
```

	Altura	Carbono	Presion	Velocidad
1	10	10	25	200
2	20	10	25	200

3	11	12	25	200
4	9	12	25	200
5	2	14	25	200
6	-1	14	25	200
7	3	10	25	250
8	5	10	25	250
9	2	12	25	250
10	5	12	25	250
11	4	14	25	250
12	7	14	25	250
13	5	10	30	200
14	9	10	30	200
15	5	12	30	200
16	4	12	30	200
17	-3	14	30	200
18	-2	14	30	200
19	-1	10	30	250
20	-3	10	30	250
21	-3	12	30	250
22	2	12	30	250
23	1	14	30	250
24	3	14	30	250

A continuación debemos transformar las tres columnas en factores para poder realizar los cálculos posteriores adecuadamente.

```
factorial$Carbono <- factor(factorial$Carbono)
factorial$Velocidad <- factor(factorial$Velocidad)
factorial$Presion <- factor(factorial$Presion)
```

Para calcular la tabla ANOVA primero hacemos uso de la función “aov” de la siguiente forma:

```
mod <- aov(Altura~ Carbono + Presion + Velocidad + Carbono*Presion +
          Carbono*Velocidad + Presion*Velocidad + Carbono*Velocidad*Presion,
          data = factorial )

summary(mod)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Carbono	2	88.08	44.04	5.683	0.018350 *
Presion	1	150.00	150.00	19.355	0.000866 ***
Velocidad	1	80.67	80.67	10.409	0.007270 **

```

Carbono:Presion      2  14.25    7.12    0.919 0.425122
Carbono:Velocidad    2 230.58   115.29   14.876 0.000564 ***
Presion:Velocidad    1   1.50    1.50    0.194 0.667799
Carbono:Presion:Velocidad 2   1.75    0.88    0.113 0.894175
Residuals           12  93.00    7.75

```

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

donde:

- **Altura:** Nombre de la columna de las observaciones
- **Carbono:** Nombre de la columna en la que está representado el primer factor
- **Presion:** Nombre de la columna en la que está representado el segundo factor
- **Velocidad:** Nombre de la columna en la que está representado el tercer factor
- **Carbono×Presion, Carbono×Velocidad, Presion×Velocidad y Carbono×VelocidadPresion** hace referencia a las distintas interacciones.
- `data= data.frame` en el que están guardados los datos

La Tabla ANOVA muestra las filas de Carbono, Presión, Velocidad, Carbono\* Presión, Carbono\* Velocidad, Presión\* Velocidad y Carbono\* Presión\* Velocidad que corresponden a la variabilidad debida a los efectos de cada uno de los factores, a las interacciones de orden dos y orden tres entre los factores. En dicha Tabla se indica que para un nivel de significación del 5% los efectos que no son significativos del modelo planteado son las interacciones entre los factores, Carbono\* Presión y Presión\* Velocidad y Carbono\* Presión\* Velocidad ya que los p-valores correspondientes a estos efectos son **0.425**, **0.668** y **0.894** mayores que el nivel de significación.

Para resolverlo suprimimos la interacción Carbono\* Presión\* Velocidad. La tabla ANOVA que corresponde a este modelo es la siguiente:

```

mod <- aov(Altura~ Carbono + Presion + Velocidad + Carbono*Presion +
           Carbono*Velocidad + Presion*Velocidad, data = factorial )
summary(mod)

```

```

              Df Sum Sq Mean Sq F value    Pr(>F)
Carbono        2  88.08   44.04    6.507 0.010038 *
Presion        1 150.00  150.00   22.164 0.000336 ***
Velocidad      1  80.67   80.67   11.919 0.003886 **
Carbono:Presion 2  14.25    7.12    1.053 0.375033
Carbono:Velocidad 2 230.58  115.29   17.035 0.000178 ***
Presion:Velocidad 1   1.50    1.50    0.222 0.645047
Residuals     14  94.75    6.77

```

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Los efectos Carbono\* Presión y Presión\* Velocidad siguen siendo no significativos.

Para resolverlo suprimimos la interacción Presión\* Velocidad. La tabla ANOVA que corresponde a este modelo es la siguiente:

```
mod <- aov(Altura~ Carbono + Presion + Velocidad + Carbono*Presion +
           Carbono*Velocidad, data = factorial )
summary(mod)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Carbono	2	88.08	44.04	6.864	0.007647	**
Presion	1	150.00	150.00	23.377	0.000218	***
Velocidad	1	80.67	80.67	12.571	0.002935	**
Carbono:Presion	2	14.25	7.12	1.110	0.355049	
Carbono:Velocidad	2	230.58	115.29	17.968	0.000104	***
Residuals	15	96.25	6.42			

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

El efecto Carbono\* Presión sigue siendo no significativo por lo tanto lo suprimimos y replanteamos el siguiente modelo matemático

```
mod <- aov(Altura~ Carbono + Presion + Velocidad + Carbono*Velocidad,
           data = factorial )
summary(mod)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Carbono	2	88.08	44.04	6.776	0.006856	**
Presion	1	150.00	150.00	23.077	0.000166	***
Velocidad	1	80.67	80.67	12.410	0.002612	**
Carbono:Velocidad	2	230.58	115.29	17.737	6.91e-05	***
Residuals	17	110.50	6.50			

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Todos los efectos de este último modelo planteado son significativos y por lo tanto es en este modelo donde vamos a realizar el estudio. Existen diferencias significativas entre los distintos porcentajes del Carbono, los dos tipos de presión, las dos velocidades de llenado y la interacción entre el porcentaje de **Carbono y la Velocidad** de llenado.



## 11 Ejercicios Guiados

### Ejercicio Guiado 1

Se realiza un estudio del contenido de azufre en cinco yacimientos de carbón. Se toman muestras aleatoriamente de cada uno de los yacimientos y se analizan. Los datos del porcentaje de azufre por muestra se indican en la tabla adjunta.

Yacimientos	Porcentaje de azufre
1	151 192 108 204 214 176 117
2	169 64 90 141 101 128 159 156
3	122 132 139 133 154 104 225 149 130
4	75 126 69 62 90 120 32 73
5	80 90 124 82 72 57 118 54 130

Para un nivel de significación del 5%.

1. ¿Se puede confirmar que el porcentaje de azufre es el mismo en los cinco yacimientos?
2. Si se rechaza la hipótesis nula que las medias de porcentaje de azufre en los cinco yacimientos es la misma, determinar que medias difieren entre sí utilizando el método de comparaciones múltiples de Tukey.
3. Estudiar las hipótesis de modelo: Homocedasticidad (Homogeneidad de las varianzas por grupo), Independencia y Normalidad.

### Solución

- 1.

El problema planteado se modeliza a través de un **diseño unifactorial totalmente aleatorizado de efectos fijos no-equilibrado**.

- **Variable respuesta: Contenido de Azufre**
- **Factor:** Tipo de yacimiento con cinco niveles. Es un factor de Efectos fijos ya que viene decidido qué niveles concretos se van a utilizar
- **Modelo no-equilibrado:** Los niveles de los factores tienen distinto número de elementos
- **Tamaño del experimento:** Número total de observaciones, en este caso 41 unidades experimentales.

```
porcentaje <- read.table("guiado1.txt", header = TRUE)
porcentaje
```

Azufre Yacimiento		
1	151	1
2	192	1
3	108	1
4	204	1
5	214	1
6	176	1
7	117	1
8	169	2
9	64	2
10	90	2
11	141	2
12	101	2
13	128	2
14	159	2
15	156	2
16	122	3
17	132	3
18	139	3
19	133	3
20	154	3
21	104	3
22	225	3
23	149	3
24	130	3
25	75	4
26	126	4
27	69	4
28	62	4
29	90	4
30	120	4
31	32	4
32	73	4
33	80	5
34	90	5
35	124	5
36	82	5
37	72	5
38	57	5
39	118	5
40	54	5
41	130	5

Debemos transformar la variable referente a los niveles del factor fijo como factor para poder hacer los cálculos de forma adecuada:

```
porcentaje$Yacimiento<-factor(porcentaje$Yacimiento)
porcentaje$Yacimiento
```

```
[1] 1 1 1 1 1 1 1 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 4 4 4 4 4 4 4 5 5 5 5 5 5
[39] 5 5 5
Levels: 1 2 3 4 5
```

Para calcular la tabla ANOVA primero hacemos uso de la función “aov” de la siguiente forma:

```
mod <- aov(Azufre ~ Yacimiento, data = porcentaje);mod
```

Call:

```
aov(formula = Azufre ~ Yacimiento, data = porcentaje)
```

Terms:

	Yacimiento	Residuals
Sum of Squares	40432.68	42639.76
Deg. of Freedom	4	36

Residual standard error: 34.41566

Estimated effects may be unbalanced

```
summary(mod)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Yacimiento	4	40433	10108	8.534	5.97e-05 ***
Residuals	36	42640	1184		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

En la Tabla ANOVA, el valor del estadístico de contraste de igualdad de medias, **F = 8.534** deja a su derecha un p-valor menor que 0.001, menor que el nivel de significación del 5%, por lo que se rechaza la Hipótesis nula de igualdad de medias. Es decir, existen diferencias significativas en el contenido medio de azufre entre los cinco yacimientos. La pregunta que nos planteamos es si el contenido de azufre es significativamente distinto en los cinco yacimientos o sólo en alguno de ellos. Para responder a esta pregunta utilizamos algún procedimiento de comparaciones múltiples. En el apartado siguiente responderemos a esta cuestión.

2.

```
mod.tukey <- TukeyHSD(mod, ordered = TRUE)
mod.tukey
```

```
Tukey multiple comparisons of means
 95% family-wise confidence level
factor levels have been ordered
```

```
Fit: aov(formula = Azufre ~ Yacimiento, data = porcentaje)
```

```
$Yacimiento
      diff      lwr      upr      p adj
5-4  8.791667 -39.217257  56.80059 0.9841364
2-4 45.125000  -4.275775  94.52577 0.0873389
3-4 62.236111  14.227188 110.24503 0.0057086
1-4 85.125000  33.990340 136.25966 0.0002709
2-5 36.333333 -11.675590  84.34226 0.2131394
3-5 53.444444   6.868947 100.01994 0.0177365
1-5 76.333333  26.542032 126.12463 0.0008288
3-2 17.111111 -30.897812  65.12003 0.8429902
1-2 40.000000 -11.134660  91.13466 0.1865081
1-3 22.888889 -26.902412  72.68019 0.6809794
```

Se comprueba que no se detectan diferencias significativas entre los yacimientos 1, 2 y 3 y entre los yacimientos 2, 4 y 5. Para ello nos fijamos en las Significaciones (mayores que 0.05) o en los límites de los intervalos. Dos medias se declaran iguales si el cero pertenece al intervalo de confianza construido para la diferencia de ellas.

3.

### Hipótesis de Homocedasticidad: Test de Barlett

```
bartlett.test(porcentaje$Azufre, porcentaje$Yacimiento)
```

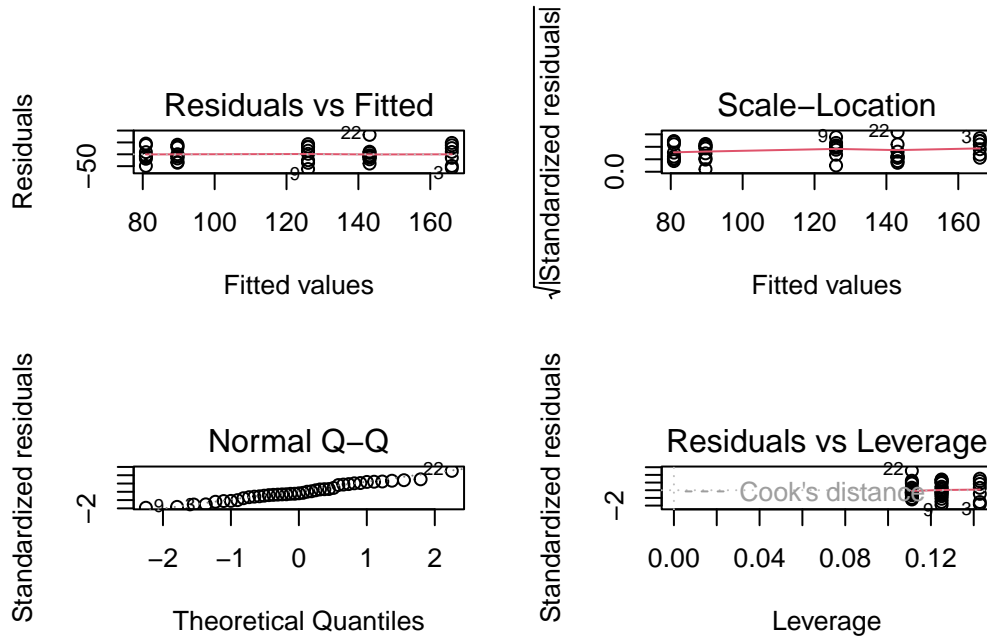
```
Bartlett test of homogeneity of variances
```

```
data:  porcentaje$Azufre and porcentaje$Yacimiento
Bartlett's K-squared = 1.2655, df = 4, p-value = 0.8672
```

La salida muestra el resultado del contraste de Barlett de igualdad de varianzas en todos los grupos. El estadístico de contraste experimental,  $B = 1.2655$ , deja a la derecha un **p-valor** = **0.8672**, que nos indica que no se debe rechazar la igualdad entre las varianzas.

**Hipótesis de Independencia:** Esta hipótesis la comprobaremos gráficamente mediante la representación de los residuos frente a los valores pronosticados por el modelo.

```
layout(matrix(c(1,2,3,4),2,2))
plot(mod)
```



En esta salida interpretamos el gráfico que se muestra en la Fila 1, Columna 1. Es decir, el gráfico el que se representan los residuos en el eje de ordenadas y los valores ajustados por el modelo en el eje de abscisas. Este gráfico no muestra ningún aspecto que haga sospechar de la hipótesis de independencia de los residuos.

La hipótesis de Normalidad la comprobaremos gráficamente y analíticamente

Gráficamente comprobaremos la normalidad mediante un histograma y el gráfico Q-Q plot

En primer lugar realizaremos el histograma, para ello

En primer lugar calculemos los residuos del modelo

```
g = mod$residuals;g
```

1	2	3	4	5	6
-15.0000000	26.0000000	-58.0000000	38.0000000	48.0000000	10.0000000
7	8	9	10	11	12
-49.0000000	43.0000000	-62.0000000	-36.0000000	15.0000000	-25.0000000
13	14	15	16	17	18
2.0000000	33.0000000	30.0000000	-21.1111111	-11.1111111	-4.1111111
19	20	21	22	23	24
-10.1111111	10.8888889	-39.1111111	81.8888889	5.8888889	-13.1111111
25	26	27	28	29	30
-5.8750000	45.1250000	-11.8750000	-18.8750000	9.1250000	39.1250000
31	32	33	34	35	36
-48.8750000	-7.8750000	-9.6666667	0.3333333	34.3333333	-7.6666667
37	38	39	40	41	
-17.6666667	-32.6666667	28.3333333	-35.6666667	40.3333333	

Calculamos la media de los residuos

```
m<-mean(g);m
```

```
[1] -3.462677e-16
```

Calculamos la desviación típica

```
std<-sqrt(var(g));std
```

```
[1] 32.64957
```

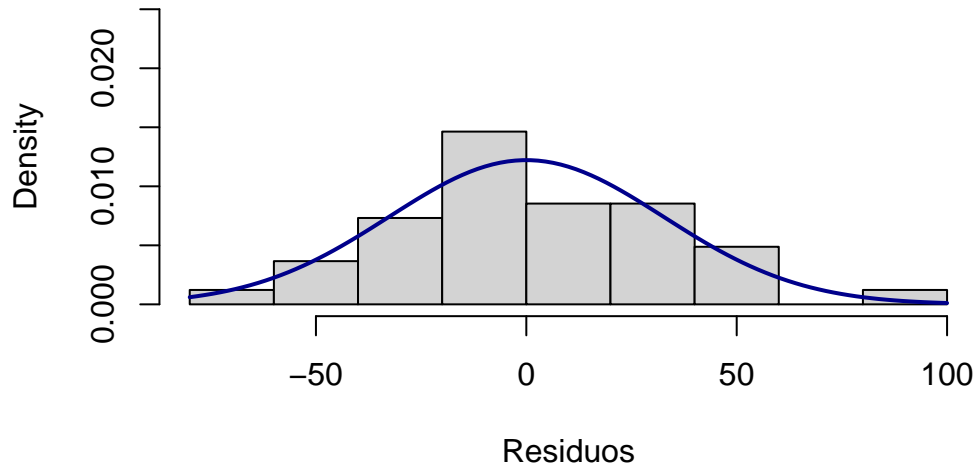
Representamos el histograma

```
hist(g, prob = TRUE, xlab = "Residuos", ylim = c(0, 0.025), main =  
      "F. dens. Normal hist. de residuos")
```

```
#Y la curva Normal sobre el Histograma
```

```
curve(dnorm(x, mean = m, sd = std), col = "darkblue", lwd = 2, add = TRUE, yaxt = "n")
```

## F. dens. Normal hist. de residuos



Anteriormente hemos realizado el gráfico Q-Q. Ambos gráficos no muestran desviación importante de la normalidad.

Analíticamente lo vamos a comprobar mediante el contraste de **Shapiro-Wilk**

```
shapiro.test(mod$residuals)
```

Shapiro-Wilk normality test

```
data:  mod$residuals  
W = 0.97902, p-value = 0.6384
```

El valor del p-valor (Sig. asintót. (bilateral)) es de **0.6384**, por lo tanto no podemos rechazar la hipótesis de normalidad.

### Ejercicio Guiado 2

Se realiza un estudio sobre el efecto del fotoperiodo y del genotipo en el periodo latente de infección del moho de cebada aislado AB3. Se obtienen cincuenta hojas de cuatro genotipos distintos. Cada grupo es infectado y posteriormente expuesto a diferente fotoperiodo. Los distintos fotoperiodos se trataron como bloques y se obtuvieron los siguientes datos de los totales para los bloques y tratamientos. La respuesta anotada es el número de días hasta la aparición de síntomas visibles.

	Fotoperiodo (horas de oscuridad por ciclo de 24 horas)				
Genotipo	0	2	4	8	16
Armelle	630	610	560	570	590
Golden	640	630	600	620	620
Promise	640	630	650	620	580
Emir	660	660	620	610	630

1. ¿Se puede afirmar que los diferentes genotipos no influyen en el número de días hasta la aparición de la infección? ¿Se puede concluir que los distintos fotoperiodos no afectan al tiempo de aparición de los síntomas de infección del moho?
2. En caso de que influyan significativamente alguno de los dos factores, extraer conclusiones utilizando el método de Duncan.
3. Estudiar las hipótesis de modelo: Homocedasticidad, Independencia y Normalidad.

### Solución

1.

En este caso se trata de un diseño en bloques completos aleatorizados. El objetivo del estudio es comparar los cuatro tipos de genotipos, por lo que se trata de un factor con cuatro niveles. Sin embargo, al realizar la medición con los distintos fotoperiodos a los que son expuestos el moho de cebada, es posible que estos influyan sobre el periodo latente de infección del moho de cebada aislado AB3. Por ello, y al no ser directamente motivo de estudio, los fotoperiodos es un factor secundario que recibe el nombre de bloque.

```
dias<-read.table("guiado2.txt", header = TRUE)
dias
```

```

      Dias Fotoperiodo Genotipo
1      630           1         1
2      640           1         2
3      640           1         3
4      660           1         4
5      610           2         1
6      630           2         2
7      630           2         3
8      660           2         4
9      560           3         1
10     600           3         2
11     650           3         3
12     620           3         4
13     570           4         1
14     620           4         2

```



15	620	4	3
16	610	4	4
17	590	5	1
18	620	5	2
19	580	5	3
20	630	5	4

A continuación debemos transformar tanto la columna de los tratamientos como la de los bloques en un factor para realizar los cálculos posteriores adecuadamente.

```
dias$Fotoperiodo = factor(dias$Fotoperiodo);dias$Fotoperiodo
```

```
[1] 1 1 1 1 2 2 2 2 3 3 3 3 4 4 4 4 5 5 5 5
Levels: 1 2 3 4 5
```

```
dias$Genotipo = factor(dias$Genotipo);dias$Genotipo
```

```
[1] 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
Levels: 1 2 3 4
```

```
# ANOVA
mod = aov(Dias ~ Fotoperiodo + Genotipo, data = dias);mod
```

Call:

```
aov(formula = Dias ~ Fotoperiodo + Genotipo, data = dias)
```

Terms:

	Fotoperiodo	Genotipo	Residuals
Sum of Squares	5030	5255	4170
Deg. of Freedom	4	3	12

Residual standard error: 18.64135

Estimated effects may be unbalanced

```
summary(mod)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Fotoperiodo	4	5030	1257.5	3.619	0.0371 *
Genotipo	3	5255	1751.7	5.041	0.0173 *
Residuals	12	4170	347.5		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

En la Tabla ANOVA, el valor del estadístico de contraste de igualdad de medias de tratamientos,  $F = 5.041$  deja a su derecha un p-valor igual a **0.017**, menor que el nivel de significación del 5%, por lo que se rechaza la Hipótesis nula de igualdad de medias de tratamientos. Es decir, existen diferencias significativas en el número de días hasta la aparición de la infección entre los cuatro genotipos.

En esta Tabla ANOVA, también se observa que el valor del estadístico de contraste de igualdad de medias de bloques,  $F = 3.619$  deja a su derecha un p-valor igual a **0.037**, menor que el nivel de significación del 5%, por lo que se rechaza la Hipótesis nula de igualdad de medias de bloques. Es decir, existen diferencias significativas en el número de días hasta la aparición de la infección entre los cinco tipos de fotoperiodos. Por lo tanto, se concluye que los niveles de ambos factores influyen de forma significativa en el número de días hasta la aparición de los síntomas de infección del moho.

2.

**Primero vamos a hacer el contraste de Duncan para los tratamientos: Genotipos**

```
library(agricolae)
duncan <-duncan.test(mod, "Genotipo", main= "Número de días con diferentes genotipos ")
duncan
```

```
$statistics
  MSerror Df  Mean      CV
  347.5 12 618.5 3.013962
```

```
$parameters
  test  name.t ntr alpha
Duncan Genotipo  4  0.05
```

```
$duncan
  Table CriticalRange
2 3.081307      25.68782
3 3.225244      26.88778
4 3.312453      27.61481
```

```
$means
  Dias      std r Min Max Q25 Q50 Q75
1  592 28.63564 5 560 630 570 590 610
2  622 14.83240 5 600 640 620 620 630
3  624 27.01851 5 580 650 620 630 640
4  636 23.02173 5 610 660 620 630 660
```

```
$comparison
NULL
```

```
$groups
  Dias groups
4  636      a
3  624      a
2  622      a
1  592      b
```

```
attr("class")
[1] "group"
```

En la tabla del factor Tipo de genotipo hay dos subconjuntos que se diferencian entre sí; el subconjunto 1 está formado por las medias del genotipo Armelle y el subconjunto 2 por las medias de los genotipos Golden, Promise y Emir. Y dentro de cada subconjunto no se aprecian diferencias significativas entre las medias. También se observa que en el genotipo Emir se produce el mayor número medio de días hasta la aparición de la infección (636) y en el genotipo Armelle se produce el menor (592).

**Segundo vamos a hacer el contraste de Duncan para los bloques: Fotoperiodos**

```
duncan1 <-duncan.test(mod,"Fotoperiodo", main=
                      "Número de días con diferentes fotoperiodos")
duncan1
```

```
$statistics
  MSerror Df  Mean      CV
  347.5 12 618.5 3.013962

$parameters
  test      name.t ntr alpha
Duncan Fotoperiodo  5  0.05
```

```
$duncan
```

```

      Table CriticalRange
2 3.081307      28.71986
3 3.225244      30.06145
4 3.312453      30.87430
5 3.370172      31.41228

$means
      Dias      std r Min Max   Q25 Q50   Q75
1 642.5 12.58306 4 630 660 637.5 640 645.0
2 632.5 20.61553 4 610 660 625.0 630 637.5
3 607.5 37.74917 4 560 650 590.0 610 627.5
4 605.0 23.80476 4 570 620 600.0 615 620.0
5 605.0 23.80476 4 580 630 587.5 605 622.5

$comparison
NULL

$groups
      Dias groups
1 642.5      a
2 632.5     ab
3 607.5      b
4 605.0      b
5 605.0      b

attr(,"class")
[1] "group"

```

En la tabla del factor Tipo de fotoperiodo hay dos subconjuntos que se diferencian entre sí; el subconjunto 1 está formado por las medias de los Fotoperiodos 0 y 2 y el subconjunto 2 por las medias de los fotoperiodos 2, 4, 8 y 16. Y dentro de cada subconjunto no se aprecian diferencias significativas entre las medias. También se observa que en el fotoperiodo 0 se produce el mayor número medio de días hasta la aparición de la infección (642.5) y en los fotoperiodos 8 y 16 se produce el menor

3.

**Estudiamos la homocedasticidad mediante el test de Barlett**

```
bartlett.test(dias$Dias, dias$Fotoperiodo)
```

```
Bartlett test of homogeneity of variances
```

```
data: dias$Dias and dias$Fotoperiodo
Bartlett's K-squared = 3.0629, df = 4, p-value = 0.5474
```

```
bartlett.test(dias$Dias, dias$Genotipo)
```

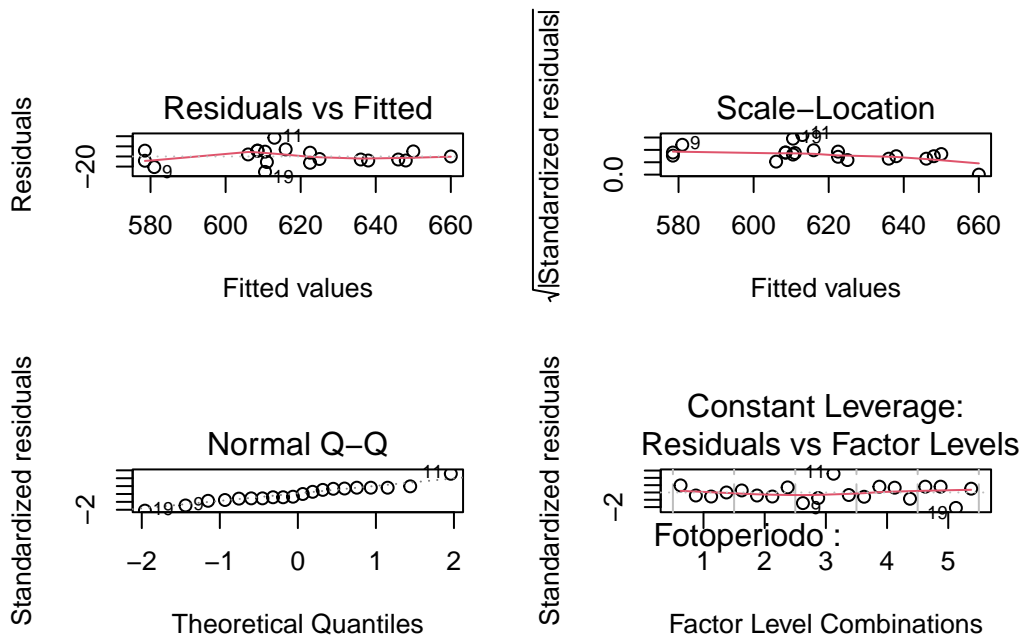
Bartlett test of homogeneity of variances

```
data: dias$Dias and dias$Genotipo
Bartlett's K-squared = 1.6252, df = 3, p-value = 0.6537
```

Las Tablas muestran los resultados del contraste de Barlett de igualdad de varianzas en todos los grupos del factor genotipo y en todos los grupos del factor Fotoperiodo. Los P-valores, 0.5474 y 0.6537 indican que indican que no se debe rechazar la igualdad entre las varianzas ni el factor genotipo ni el factor fotoperiodo.

**Estudiamos la independencia gráficamente**

```
layout(matrix(c(1,2,3,4),2,2))
plot(mod)
```



## Estudiamos la Normalidad gráficamente mediante el gráfico probabilístico normal y analíticamente mediante el contraste de Shapiro-Wilk

Observamos el gráfico que se muestra en la Fila 2, Columna1. Es decir, el gráfico el que se representan los residuos estandarizados en el eje de ordenadas y cuantiles teóricos en el eje de abscisas. En dicho gráfico se aprecian desviaciones a la normalidad, pero el contraste ANOVA es robusto frente a desviaciones pequeñas de la normalidad. Realizaremos a continuación el contraste de Shapiro-Wilk para comprobar analíticamente la normalidad de los residuos.

```
shapiro.test(mod$residuals)
```

Shapiro-Wilk normality test

```
data:  mod$residuals  
W = 0.95316, p-value = 0.4176
```

El valor del p-valor es de 0.4176, no pudiéndose rechazar la hipótesis de normalidad.

### Ejercicio Guiado 3

Se realiza un estudio para determinar el efecto del nivel del agua y del tipo de planta sobre la longitud global del tallo de las plantas de guisantes. Para ello, se utilizan tres niveles de agua (bajo, medio y alto) y dos tipos de plantas (sin hojas y convencional). Se dispone para el estudio de dieciocho plantas sin hojas y dieciocho plantas convencionales. Se dividen aleatoriamente los dos tipos de plantas en tres subgrupos y después se asignan los niveles de agua aleatoriamente a los dos grupos de plantas. Los datos sobre la longitud del tallo de los guisantes (en centímetros) se muestran en la siguiente tabla:

Para un nivel de significación del 5%.

1. ¿Se puede afirmar que los distintos niveles de agua influyen en la longitud del tallo de los guisantes? ¿Y el tipo de planta?
2. ¿La efectividad del nivel del agua es la misma para los dos tipos de plantas?
3. Estudia, utilizando el método de Newman-Keuls, qué nivel de agua es más efectivo.

### Solución

- 1.

El problema planteado se modeliza a través de un 'diseño de dos factores con replicación.

```
factorial<-read.table("guiado3.txt", header = TRUE)  
factorial
```

	Nivel del agua		
Tipo de planta	Bajo	Medio	Alto
Sin hojas	69.50	96.10	121.00
	69.00	102.30	122.90
	75.00	107.50	123.10
	70.00	103.60	125.70
	74.40	100.70	125.20
	75.00	101.80	120.10
Convencional	71.10	81.00	101.10
	69.20	85.80	103.20
	70.40	86.00	106.10
	73.20	87.50	109.70
	71.20	88.10	110.00
	70.90	87.60	99.00

	Longitud_tallo	Nivel_agua	Tipo_planta
1	69.5	1	1
2	69.0	1	1
3	75.0	1	1
4	70.0	1	1
5	74.4	1	1
6	75.0	1	1
7	96.1	2	1
8	102.3	2	1
9	107.5	2	1
10	103.6	2	1
11	100.7	2	1
12	101.8	2	1
13	121.0	3	1
14	122.9	3	1
15	123.1	3	1
16	125.7	3	1
17	125.2	3	1
18	120.1	3	1
19	71.1	1	2
20	69.2	1	2

21	70.4	1	2
22	73.2	1	2
23	71.2	1	2
24	70.9	1	2
25	81.0	2	2
26	85.8	2	2
27	86.0	2	2
28	87.5	2	2
29	88.1	2	2
30	87.6	2	2
31	101.1	3	2
32	103.2	3	2
33	106.1	3	2
34	109.7	3	2
35	110.0	3	2
36	99.0	3	2

A continuación debemos transformar todas las columnas que contienen a los factores en un factor para poder realizar los cálculos posteriores adecuadamente.

```
factorial$agua <- factor(factorial$Nivel_agua);factorial$agua
```

```
[1] 1 1 1 1 1 1 2 2 2 2 2 2 3 3 3 3 3 3 1 1 1 1 1 1 2 2 2 2 2 2 3 3 3 3 3 3
Levels: 1 2 3
```

```
factorial$planta <- factor(factorial$Tipo_planta);factorial$planta
```

```
[1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
Levels: 1 2
```

Para responder a este apartado debe resolverse el contraste de igualdad de medias para el factor A:  $H_0 = \mu_1 = \mu_2 = \mu_3 = \mu$  vs  $H_1 = \mu_i \neq \mu_j$  para algún  $i \neq j$

y para el factor B:  $H_0 = \mu_1 = \mu_2 = \mu$  vs  $H_1 = \mu_i \neq \mu_j$  para algún  $i \neq j$

```
mod = aov(Longitud_tallo ~ agua* planta , data = factorial )
mod
```



Call:

```
aov(formula = Longitud_tallo ~ agua * planta, data = factorial)
```

Terms:

	agua	planta	agua:planta	Residuals
Sum of Squares	10773.635	1246.090	514.145	282.250
Deg. of Freedom	2	1	2	30

Residual standard error: 3.067301

Estimated effects may be unbalanced

```
summary(mod)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
agua	2	10774	5387	572.56	< 2e-16 ***
planta	1	1246	1246	132.44	1.58e-12 ***
agua:planta	2	514	257	27.32	1.75e-07 ***
Residuals	30	282	9		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

- El valor del estadístico de contraste de igualdad de medias del factor Nivel\_agua, **F= 572.56** deja a su derecha un p-valor menor que **0.001**, menor que el nivel de significación del 5%, por lo que se rechaza la Hipótesis nula de igualdad de medias de los niveles del factor Nivel\_agua. Es decir, existen diferencias significativas en la longitud del tallo de guisantes dependiendo del nivel del agua.
- El valor del estadístico de contraste de igualdad de medias del factor Tipo\_planta, **F= 132.445** deja a su derecha un p-valor menor que **0.001**, menor que el nivel de significación del 5%, por lo que se rechaza la Hipótesis nula de igualdad de medias del factor Tipo\_planta. Es decir, el tipo de planta afecta significativamente a la longitud del tallo de guisantes.

2.

Para responder a esta pregunta, realizamos el contraste de hipótesis sobre la interacción de los dos factores  $H0 = (\tau\beta)_{ij} = 0$  (no existe interacción) vs  $H1 = (\tau\beta)_{ij} \neq 0$  (existe interacción).

En la Tabla ANOVA mostrada anteriormente, el valor del estadístico de contraste de la interacción de los dos factores, **F= 27.32** deja a su derecha un p-valor menor que **0.001**, menor que el nivel de significación del 5%, por lo que se rechaza la Hipótesis nula de no interacción entre los factores. Por lo tanto la efectividad del nivel de agua no es la misma para los dos tipos de plantas. Es decir, puede ocurrir que un nivel de agua influya en el crecimiento de la longitud del tallo con un tipo de planta pero no con el otro o influya de distinta forma.

3.

```
library(agricolae)
contraste <- SNK.test(mod,"agua", console=TRUE,
                      main="Contraste de Newman-Keuls para el factor nivel del agua")
```

Study: Contraste de Newman-Keuls para el factor nivel del agua

Student Newman Keuls Test  
for Longitud\_tallo

Mean Square Error: 9.408333

agua, means

	Longitud_tallo	std	r	Min	Max
1	71.575	2.243019	12	69	75.0
2	94.000	8.901992	12	81	107.5
3	113.925	10.069949	12	99	125.7

Alpha: 0.05 ; DF Error: 30

Critical Range  
2 3  
2.557375 3.087063

Means with the same letter are not significantly different.

	Longitud_tallo	groups
3	113.925	a
2	94.000	b
1	71.575	c

En la tabla se muestran los subgrupos formados de medias iguales al utilizar el método de Newman-Keuls. Hay tres subconjuntos que se diferencian entre sí y cada subconjunto está formado por un solo nivel de agua. También se observa que con el nivel de agua alto se produce la mayor longitud del tallo de guisantes, **113.925 cm**, y con el nivel Bajo se produce el menor **71.575 cm**.

## 12 Ejercicios Propuestos

### Ejercicio Propuesto 1

La convección es una forma de transferencia de calor por los fluidos debido a sus variaciones de densidad por la temperatura; las partes calientes ascienden y las frías descienden formando las corrientes de convección que hacen uniforme la temperatura del fluido. Se ha realizado un experimento para determinar las modificaciones de la densidad de fluido al elevar la temperatura en una determinada zona. Los resultados obtenidos han sido los siguientes:

Temperatura	Densidad				
100	21.8	21.9	21.7	21.6	21.7
125	21.7	21.4	21.5	21.4	
150	21.9	21.8	21.8	21.6	21.5
175	21.9	22.1	21.85	21.9	

Responder a las siguientes cuestiones:

1. ¿Afecta la temperatura a la densidad del fluido?
2. Determinar qué temperaturas producen modificaciones significativas en la densidad media del fluido.
3. Estudiar las hipótesis del modelo: Homocedasticidad, independencia y normalidad.
4. Se puede afirmar que las temperaturas de 100 y 125 producen menos densidades de fluido en promedio que las temperaturas de 150 y 175.

### Solución

El problema planteado se modeliza a través de un diseño unifactorial totalmente aleatorizado de efectos fijos no-equilibrado.

1.

```
propuesto1 <- read.table("s9_propuesto1.txt", header = TRUE)
propuesto1
```

	Densidad	Temperatura
1	21.80	100
2	21.90	100
3	21.70	100
4	21.60	100
5	21.70	100
6	21.70	125
7	21.40	125

8	21.50	125
9	21.40	125
10	21.90	150
11	21.80	150
12	21.80	150
13	21.60	150
14	21.50	150
15	21.90	175
16	22.10	175
17	21.85	175
18	21.90	175

Transformando la variable referente a los niveles del factor fijo como factor

```
propuesto1$Temperatura <- factor(propuesto1$Temperatura)
propuesto1$Temperatura
```

```
[1] 100 100 100 100 100 125 125 125 125 150 150 150 150 150 175 175 175 175
Levels: 100 125 150 175
```

Calculando el ANOVA

```
mod <- aov(Densidad ~ Temperatura, data = propuesto1)
mod
```

Call:

```
aov(formula = Densidad ~ Temperatura, data = propuesto1)
```

Terms:

	Temperatura	Residuals
Sum of Squares	0.384375	0.256875
Deg. of Freedom	3	14

Residual standard error: 0.1354556

Estimated effects may be unbalanced

```
summary(mod)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Temperatura	3	0.3844	0.12813	6.983	0.00419 **
Residuals	14	0.2569	0.01835		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

El valor de F experimental 6.983, deja a su derecha un p-valor = 0.00419 inferior a 0.05, por lo que se rechaza la hipótesis nula de igualdad de medias. Concluyendo que existen diferencias significativas en la densidad del fluido en función de la modificación de la temperatura

2.

```
mod.tukey <- TukeyHSD(mod, ordered = TRUE);mod.tukey
```

```
Tukey multiple comparisons of means
 95% family-wise confidence level
factor levels have been ordered
```

```
Fit: aov(formula = Densidad ~ Temperatura, data = propuesto1)
```

```
$Temperatura
      diff      lwr      upr      p adj
150-125 0.2200 -0.04410917 0.4841092 0.1184044
100-125 0.2400 -0.02410917 0.5041092 0.0807121
175-125 0.4375  0.15910449 0.7158955 0.0021907
100-150 0.0200 -0.22900452 0.2690045 0.9953076
175-150 0.2175 -0.04660917 0.4816092 0.1240769
175-100 0.1975 -0.06660917 0.4616092 0.1785203
```

Contraste de comparaciones múltiples de Duncan

```
duncan <-duncan.test(mod, "Temperatura" ,
  main= " Modificaciones de la densidad de fluido al elevar la temperatura ")
```

```
duncan
```

```
$statistics
      MSerror Df      Mean      CV
0.01834821 14 21.725 0.6235009
```

```
$parameters
```

```

      test      name.t ntr alpha
Duncan Temperatura    4  0.05

$duncan
NULL

$means
      Densidad      std r   Min  Max      Q25   Q50   Q75
100  21.7400 0.1140175 5 21.60 21.9 21.7000 21.70 21.80
125  21.5000 0.1414214 4 21.40 21.7 21.4000 21.45 21.55
150  21.7200 0.1643168 5 21.50 21.9 21.6000 21.80 21.80
175  21.9375 0.1108678 4 21.85 22.1 21.8875 21.90 21.95

$comparison
NULL

$groups
      Densidad groups
175  21.9375      a
100  21.7400      b
150  21.7200      b
125  21.5000      c

attr(,"class")
[1] "group"

```

No existe suficiente evidencia para rechazar la hipótesis.

3.

### Hipótesis de HOMOCEDASTICIDAD

```
bartlett.test(propuesto1$Densidad, propuesto1$Temperatura)
```

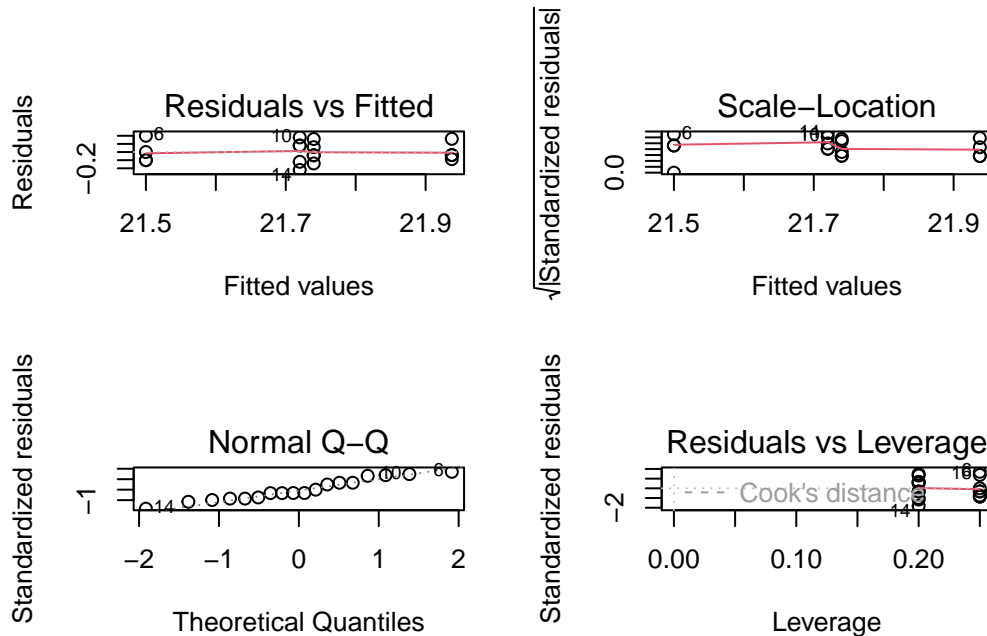
Bartlett test of homogeneity of variances

```
data: propuesto1$Densidad and propuesto1$Temperatura
Bartlett's K-squared = 0.69212, df = 3, p-value = 0.8751
```

B= 0.69212, y un p-valor = 0.8751, que nos indica que no se existe suficiente evidencia para rechazar la igualdad entre las varianzas entre las densidades registradas para las diferentes temperaturas

## Hipótesis de INDEPENDENCIA

```
layout(matrix(c(1,2,3,4),2,2))
plot(mod)
```



. Para comprobar que se satisface el supuesto de independencia entre los residuos, observamos el primer gráfico que se representan los residuos en el eje de ordenadas y los valores pronosticados en el eje de abscisas. La presencia de alguna tendencia en el gráfico puede indicar la alteración de dicha hipótesis. En este gráfico no se observa ninguna tendencia concreta lo que muestra la no existencia de relación de dependencia.

## Hipótesis de NORMALIDAD

```
# Calculo de los residuos del modelo
g = mod$residuals

# Calculamos la media de los residuos
m <- mean(g);m
```

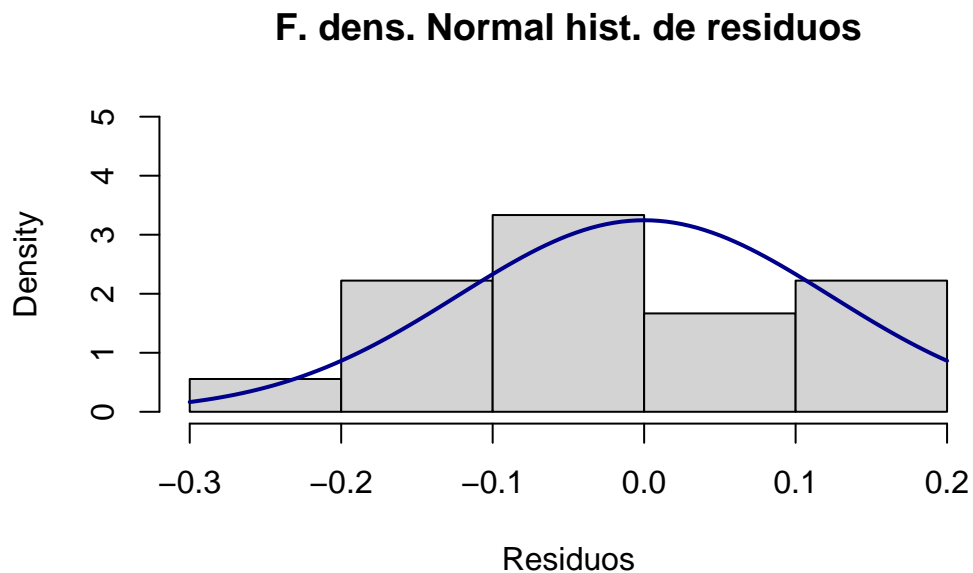
```
[1] -6.717912e-19
```

```
# Calculamos la desviación típica
std <- sqrt(var(g))
std
```

```
[1] 0.1229239
```

```
# Representamos el histograma
hist(g, prob = TRUE, xlab = "Residuos", ylim = c(0, 5),
     main = "F. dens. Normal hist. de residuos")

# Curva normal sobre el histograma
curve(dnorm(x, mean = m, sd = std), col = "darkblue", lwd = 2, add = TRUE, yaxt = "n")
```



El gráfico Q-Q, no muestra desviaciones importantes a la normalidad

Comprobando la hipótesis de normalidad mediante el contraste de Shapiro-Wilk

```
shapiro.test(mod$residuals)
```

```
Shapiro-Wilk normality test
```



```
data: mod$residuals
W = 0.94639, p-value = 0.3712
```

El valor del p-valor es de 0.3712, por lo tanto no existe suficiente evidencia para rechazar la hipótesis de normalidad de los residuos

4.

El test de Duncan nos muestra:

```
duncan
```

```
$statistics
```

MSerror	Df	Mean	CV
0.01834821	14	21.725	0.6235009

```
$parameters
```

test	name.t	ntr	alpha
Duncan	Temperatura	4	0.05

```
$duncan
```

```
NULL
```

```
$means
```

	Densidad	std r	Min	Max	Q25	Q50	Q75	
100	21.7400	0.1140175	5	21.60	21.9	21.7000	21.70	21.80
125	21.5000	0.1414214	4	21.40	21.7	21.4000	21.45	21.55
150	21.7200	0.1643168	5	21.50	21.9	21.6000	21.80	21.80
175	21.9375	0.1108678	4	21.85	22.1	21.8875	21.90	21.95

```
$comparison
```

```
NULL
```

```
$groups
```

	Densidad	groups
175	21.9375	a
100	21.7400	b
150	21.7200	b
125	21.5000	c

```
attr(,"class")
```

```
[1] "group"
```

Se muestra en la tabla que las temperaturas de 150 y 125 producen menos densidades de fluido en promedio que las temperaturas de 100 y 175